Carbohydrate-induced gastrointestinal symptoms: development and validation of a test-specific symptom questionnaire for an adult population, the adult Carbohydrate Perception Questionnaire

Johann Hammer[a], Marc Sonyi[b], Katrin M. Engeßer[a], Guntram Riedl[a], Stefan Luong[a] and Heinz F. Hammer[b]

Objectives Carbohydrate intolerances may affect a majority of the worlds-population but there is no validated, test-specific assessment of carbohydrate-induced symptoms during breath tests. We aimed to develop and validate a questionnaire for evaluation and quantification of carbohydrate intolerance.

Methods A visual analog scale-questionnaire with five complaints (pain, nausea, bloating, flatulence, and diarrhea) was designed. The time frame of symptoms was 'current' (for baseline symptoms) and 'since filling out the last questionnaire'. Validity was determined in focus-group style interviews and during breath tests in an original ($n = 342$) and follow-up patient groups ($n = 338$).

Results The questionnaire had good face validity, content validity ratio according to Lawshe was 1. Intraclass correlation coefficients ($n = 195$; 30-min' interval) demonstrated excellent reliability ($P < 0.001$), Cohen's d (measure of effect size) was small (≤0.19 for each symptom). Convergent and discriminant validity were supported against patient interviews. Questionnaire-derived results highly correlated with a medical interview ($P < 0.001$; $n = 338$). Responsiveness to change was verified during breath tests despite small effect sizes (≤0.32). Additional cross-validation and external validation studies (follow-up in-house: $n = 182$; external: $n = 156$) demonstrated generalizability and identified relevant numbers of patients in whom there was no co-occurrence of carbohydrate malabsorption and intolerance.

Conclusions The adult Carbohydrate Perception Questionnaire is a valid instrument for the assessment of gastrointestinal symptoms after carbohydrate ingestion with excellent psychometric properties. It allows standardized, test-specific diagnosis of carbohydrate intolerance and evaluation of the relation between malabsorption and intolerance. It shall be useful for future studies on treatment of carbohydrate intolerance. Eur J Gastroenterol Hepatol XXX: 00–00

Introduction

Ingestion of carbohydrates in amounts that exceed small intestinal absorptive capacities can cause abdominal symptoms. Malabsorbed carbohydrates reach the colon where bacterial metabolism converts them into gases and short-chain fatty acids [1,2] which are partly responsible for symptoms of carbohydrate malabsorption [3,4], although carbohydrate-induced symptoms can also arise without detectable malabsorption [5,6].

Breath tests are commonly used tools for the diagnosis of carbohydrate malabsorption by measuring hydrogen in exhaled air after ingestion of provocative doses of carbohydrates such as lactose or fructose. These tests are inexpensive, simple, well tolerated, and widely used. However, studies in which unvalidated symptom assessments during breath tests were used have demonstrated a discrepancy between malabsorption and development of symptoms, that is, intolerance [5,7]. Therefore, the clinical relevance of malabsorption without symptom assessment is disputed [8–12], with some authors proposing a shift of the clinical focus to evaluation of symptoms [10]. In fact, the decision on starting treatment with diet or enzyme replacement shall focus on carbohydrate-intolerance, and not on malabsorption [13].

A prerequisite for diagnosing carbohydrate intolerance is a valid, standardized evaluation of symptoms. In the absence of unbiased symptom assessment, breath hydrogen testing has been considered of limited value in guiding dietary treatment as dietary restriction has let to conflicting results [11,14]. While some guidelines recommend the evaluation of symptoms during carbohydrate challenge tests [12,15] others recognize the current lack of effective symptom assessment in the absence of validated scales [16].

Here, we describe the development and validation of a self-administered symptom measurement questionnaire to assess the severity and the type of abdominal symptoms after an oral carbohydrate load, the adult Carbohydrate Perception Questionnaire (aCPQ). The questionnaire underwent a rigorous development and validation process

European Journal of Gastroenterology & Hepatology 2020, XXX:00–00

Keywords: breath test, carbohydrate, fructose, hypersensitivity, intolerance, lactose

[a]Department of Gastroenterology and Hepatology, University Hospital of Internal Medicine 3; Medical University of Vienna and [b]Department of Gastroenterology and Hepatology, Medical University Graz, Graz, Austria

Correspondence to ao. Univ. Prof. Dr. Johann Hammer, Abteilung für Gastroenterologie und Hepatologie, Universitätsklinik für Innere Medizin 3, Währinger Gürtel 18 – 20, A-1090 Vienna, Austria

Tel: +0043 1 40400 47410; e-mail: Johann.Hammer@meduniwien.ac.at

Received 8 March 2020 **Accepted** 7 June 2020

including the implementation of the scale in large population cohorts for internal and external validation. Our aim was to overcome the lack of availability of standardized and validated symptom assessment during carbohydrate breath tests [17] to minimize bias and to develop a standard tool for the diagnosis of carbohydrate intolerance, which may be used in studies evaluating pathophysiology and treatment of patients with carbohydrate intolerance.

Methods

Questionnaire development

After literature search and initial focus group-style interviews given to patients who underwent breath hydrogen (H_2) testing and to five physicians and three technicians experienced in breath testing, five relevant complaints were identified and a questionnaire was constructed.

Adult Carbohydrate Perception Questionnaire

The symptoms evaluated were pain, nausea, bloating, flatulence, and diarrhea in German language. Responses were given on a 100-mm visual analog scale (VAS) with the words 'none' at the left and 'very severe' at the right. The time frame of symptoms was given as 'current' (for baseline symptoms) and 'since filling out the last questionnaire' (for assessment after ingestion of the carbohydrate).

Study population

All patients studied underwent carbohydrate H_2-breath testing as part of the diagnostic workup of functional gastrointestinal disorders after ruling out organic disease as clinically indicated. Some patients had tests with two different carbohydrates.

The initial cohort (Vienna-original group) consisted of 342 consecutive adult outpatients (46.8 ± 0.9 years; 211 female fructose: $n = 147$, lactose: $n = 195$) from the Vienna study center. The trial ran from March 2017 to June 2018. Cross-validity and external validity were determined in two follow-up populations for out-of-sample testing after initial implementation of the questionnaire. Cross-validity was assessed in a follow-up cohort ($n = 182$; 122 female; age: 43.8 ± 1.3 years; fructose: $n = 98$; lactose: $n = 84$) from the same institution as the validation set (Vienna-cross group). External validity was determined in a group from the other study center (Graz-external group, $n = 156$; 94 female; age: 40.2 ± 1.5 years; fructose: $n = 98$; lactose: $n = 84$).

The study was approved by the institutional ethics committees of the Medical Universities Vienna (EKNr2049/2017) and Graz (EKNr29-467ex16/17).

Breath tests and scale administration

Breath tests were performed by experienced technical assistants. Patients received standardized instructions before the test. End-expiratory breath samples were collected and analyzed for H_2 using GMI-H2-Analyzer (Stimotron medical devices, Hamburg, Germany) in Vienna and GastroCH_4ECK Gastrolyzer (Bedfont Scient. Ltd. Dr. Lahner, Anif, Austria) in Graz. The first alveolar breath sample was collected at baseline before 25 g fructose or 50 g lactose (Kwizda Pharma, Vienna, Austria) was ingested. Breath samples were analyzed 30, 60, 120, and 180 min after carbohydrate ingestion. An increase in the exhalation of H_2 ≥20 parts per million (ppm) over baseline was considered positive (i.e. malabsorption).

Patients were required to complete the aCPQ before carbohydrate ingestion (baseline) and thereafter, concurrent with collection of breath samples. Patients were asked to complete additional questionnaires 3 and 6 h after the breath test was terminated and the patients had resumed their daily routine. These two additional questionnaires were delivered by mail or at the next visit. These additional questionnaires were provided by 220 patients (Vienna-original group). A diagnosis of intolerance was a priori defined as an increase of ≥20 mm over baseline of at least one symptom assessed by the aCPQ during the 3 h of breath testing.

Questionnaire validation

Reliability

In 195 patients who underwent lactose challenge, test–retest reliability of the measure was assessed using the first (filled out before lactose ingestion) and second questionnaire (filled out 30 min after lactose ingestion). Patients undergoing fructose challenge were excluded from test–retest analysis, as symptoms arise earlier after fructose as compared to lactose and may be present 30 min after fructose ingestion [5,7]. Test–retest reliability was established, first, by evaluating the average within-patient change of each symptom score over the 30-min' interval (statistical inference via the Wilcoxon signed-ranks test), and second, by type C intraclass correlation coefficients [ICC (3,1) (two-way mixed effects, consistency, single rater/measurement for each item)]. The between-measure variance was excluded from the denominator variance. Cohen's d was calculated as a measure of effect size.

Construct validity

A total of 80 patients (mean age ± SEM: 47.7 ± 1.8 years; 47 female) undergoing the carbohydrate breath test consented to an interview during the breath test for detailed assessment of face and content validity of the aCPQ. The interviewers were blinded as to the result of the questionnaire and conducted the standard interviews after the 120-min breath sample. For the assessment of face validity, patients were asked five questions that could be answered on a five-point ordinal scale (Table 1). For content validity patients were asked to rate two statements by yes/indifferent/no (Table 2). In an additional open question, the patients were invited to note whether there are important complaints that were not captured in the questionnaire.

Additionally, five experienced physicians were asked whether they considered the questionnaire to cover all essential symptoms (yes, moderate, and no) and the content validity ratio according to Lawshe [18] was calculated.

An end-of-breath-test interview was performed in the Vienna-original group: patients were asked at the end of the breath test to indicate whether or not they had experienced the following symptoms during the breath test: pain, bloating, flatulence, and diarrhea. The interviewers were blinded as to the result of the questionnaire. Pearson

Table 1. Face validity (80 patients)

	Yes	Mostly yes	Moderate	Mostly no	No
Is the questionnaire easy to understand?	69 (86.3%)	10 (12.5%)	1 (1.3%)	0 (0%)	0 (0%)
Is it unambiguous and clear what is meant by the questions?	73 (91.3%)	7 (8.8%)	0 (0%)	0 (0%)	0 (0%)
Is it easy to answer the questions for complaints?	65 (81.3%)	13 (16.3%)	1 (1.3%)	1 (1.3%)	0 (0%)
Do you think that the questions ask for all relevant complaints?	44 (55.0%)	27 (33.8%)	8 (10.0%)	1 (1.3%)	0 (0%)
	Very easy	Easy	Moderate	Difficult	Very difficult
How difficult is it for you to grade the severity of symptoms?	61 (76.3%)	17 (21.3%)	2 (2.5%)	0 (0%)	0 (0%)

Table 2. Content validity (80 patients)

	Yes	Indifferent	No
The essential symptoms are considered in the questionnaire	74 (92.5%)	6 (7.5%)	0 (0%)
I consider the questions useful to communicate my symptoms	70 (87.5%)	10 (12.5%)	0 (0%)

correlation analysis was performed and correlation coefficients between individual aCPQ score against the post-test interview were used to determine convergent and discriminant validity. In order to determine concurrent validity, the interview was interpreted to be positive, that is, to show sensitivity to (equivalent to intolerance for) the test carbohydrate, if at least one (post-test interview 1+) or at least 2 (post-test interview 2+) symptoms were reported to have arisen during the test. Phi coefficient was calculated to evaluate the significance of correlation between the dichotomous variables, post-test interview, and aCPQ.

Cronbach's alpha was calculated as a measure of internal consistency after calculating the sum of each item score over the observation period.

Responsiveness

To assess responsiveness to change after a carbohydrate challenge in the Vienna-original group, a one-way-repeated analysis of variance- test (ANOVA) (Wilks–Lambda test) was used to explore for difference in questionnaire responses across subsequent tests and eta squared (η^2) was reported as a measure of overall effect size. As a second analysis step, symptom scores were compared between each subsequent questionnaire (min 0 vs. 30; min 30 vs. 60; etc.) using Wilcoxon signed-rank tests. The whole test period (min 0–540) was examined, each individual symptom and a global symptom score were assessed. The global symptom score was calculated for each participant by summing each individual symptom score at different time points. Only patients who reported symptoms during the breath test at the post-test interview 2+ were included in this analysis. Because each symptom arises at different points in time and symptoms after fructose manifest earlier than after lactose [5,7], we expected small effect sizes determined by this across group test statistics. Individual effect sizes from one time-point to the next were determined as Cohen's d-value.

Generalizability and external validation

To test for generalizability of the validation data, the proportion of patients with intolerance among malabsorbers and the proportion of patients with malabsorption among intolerant patients in the Vienna-cross and the Graz-external groups were determined and compared to the Vienna-original group by logistic regression analysis before and after correcting for age, sex, and the carbohydrate used.

Statistical analysis

Statistical analysis was performed using SPSS 24 (IBM SPSS Statistics for Windows, Version 24.0, released 2016, IBM Corp; Armonk, New York, USA). Data are given as mean ± SEM or median (25th/75th percentile) as appropriate. A *P*-value of <0.05 was considered significant.

Results

Test–retest reliability

When 195 patients were given the aCPQ twice, their paired scores for the five symptom-items did not change significantly (*P*-values >0.05 for all items,). Moreover, correlation of items was highly significant. These results are supportive of excellent agreement between occasions [19]. Values for Cohen's d, in which values <0.4 are considered small, were well inside the small range (Table 3). In summary, these data support test–retest reliability.

Construct validity

Face validity

The majority of patients interviewed ($n = 80$) perceived the questionnaire as unambiguous and clear, easy to answer, and relevant (Table 1). The questionnaire, therefore, has strong face validity in that it was simple, easy to understand and brief.

Content validity

The majority of interviewed patients ($n = 80$) considered the content of the questionnaire to be useful and complete (Table 2). The following additional symptoms were mentioned for possible relevance in a questionnaire: singultus (1 patient), borborygmi (1), abdominal stinging (1), constipation (1), headache (2), fatigue (1), nasal mucosal irritation (1), and assessment of the quality of pain (1). Content validity ratio according to Lawshe equaled 1, which is excellent content validity.

The intercorrelations between items on the aCPQ were low, suggesting a lack of redundancy of items (Table 4).

Table 3. Test–retest reliability for the adult Carbohydrate Perception Questionnaire-items ($n = 195$): paired symptom item scores obtained before and 30 min after lactose

Symptom-item	Mean change[a]	SD[b]	P[c]	ICC[d]	P[e]	Effect size (Cohen's d)
Pain	0.05	15.16	0.51	0.85	<0.001	–0.003
Nausea	2.68	19.96	0.42	0.81	<0.001	–0.13
Meteorism	0.79	19.98	0.86	0.83	<0.001	–0.04
Flatulence	3.24	20.99	0.12	0.83	<0.001	0.15
Diarrhea	4.53	24.30	0.11	0.77	<0.001	0.19

[a,b]Average within patient change (in mm, out of a maximal possible change of 100 in the VAS scale) and SD.
[c]*P* value from Wilcoxon signed-ranks test.
[d]ICC, intra-class correlation coefficient
[e]*P* value from Pearson correlation.

Table 4. Intercorrelations between items on the adult Carbohydrate Perception Questionnaire (above) and correlation between adult Carbohydrate Perception Questionnaire and the post-test interview (below)

	aCPQ pain (yes/no)	aCPQ nausea (yes/no)	aCPQ bloating (yes/no)	aCPQ flatulence (yes/no)	aCPQ diarrhea (yes/no)
aCPQ pain (yes/no)	1.000	0.261	0.567	0.321	0.184
aCPQ nausea (yes/no)		1.000	0.328	0.209	–0.081
aCPQ bloating (yes/no)			1.000	0.409	0.104
aCPQ flatulence (yes/no)				1.000	0.097
aCPQ diarrhea (yes/no)					1.000
Post-test evaluation pain	0.341	0.130	0.222	0.251	0.059
Post-test evaluation bloating	0.321	0.186	0.334	0.282	0.059
Post-test evaluation flatulence	0.288	0.201	0.281	0.344	0.089
Post-test evaluation diarrhea	0.280	0.108	0.222	0.234	0.496

aCPQ, adult Carbohydrate Perception Questionnaire.

Table 5. Concurrent validity (342 patients)

	Positive	Negative	Missing data	Sum
Post-test interview 1+				
aCPQ				
Positive	134	35	1	170
Negative	65	104	3	172
Sum	199	139	4	342
Post-test interview 2+				
aCPQ				
Positive	123	46	1	170
Negative	36	133	3	172
Sum	159	179	4	342

Correlation between the aCPQ and an interview after the breath test. The interview was positive, equivalent to intolerance for the test solution if at least one (post-test interview 1+) or at least 2 (post-test interview 2+) symptoms were reported to have arisen during the test.
aCPQ, adult Carbohydrate Perception Questionnaire.

Convergent and discriminant validity

All the symptoms in the aCPQ correlated highest with the same symptom in the post-test interview, supporting convergent and discriminant validity (Table 4). The post-test interview correlated significantly with the result of the aCPQ (phi statistic $P < 0.001$; concurrent validity); this was independent of whether at least one or at least two symptoms were considered as positive (Table 5).

Cronbach's alpha was 0.85, indicating good internal consistency.

Responsiveness to change

The ANOVA test statistic was <0.001 for global symptom scores indicating one or more significant changes of symptom severity during the observation; the effect size η^2 was 0.22 (Table 6). Likewise, ANOVA was significant for changes of severity for each individual symptom (pain $P = 0.002$; $\eta^2 = 021$; nausea $P = 0.01$; $\eta^2 = 014$; meteorism $P = 0.007$; $\eta^2 = 0.17$; flatulence $P < 0.001$; $\eta^2 = 0.22$; diarrhea $P < 0.001$; $\eta^2 = 0.19$). When symptom scores were compared between each subsequent questionnaire (min 0 vs. 30; min 30 vs. 60; etc.), significant changes were observed at different time points for each individual symptom.

Generalizability and external validation

In the 'Vienna-original' group, 42.7% ($n = 146$ of 342 patients) were diagnosed with malabsorption (positive breath-H_2 test) and 52.4% ($n = 179$) with intolerance (positive aCPQ); 28.4% ($n = 97$) had both malabsorption AND intolerance, 24.0% ($n = 82$) had intolerance only, and 14.3% ($n = 49$) had only malabsorption but no intolerance (Table 7).

The 'Vienna-cross' group comprised more fructose tests (54%; $P = 0.02$) and was younger (43 ± 1.3 years; $P = 0.05$) than the 'Vienna-original' group. The percentage of patients diagnosed with intolerance AND malabsorption, only intolerance or only malabsorption was comparable in both 'Vienna' groups before and after correcting for age, sex, and carbohydrate tested (NS).

The 'Graz-external' group was younger (40 ± 1.5 years; $P < 0.001$) than the 'Vienna-original' group. The percentage of patients diagnosed with intolerance AND malabsorption, only intolerance or only malabsorption was comparable in 'Graz-external' and 'Vienna-original' before and after correcting for age, sex, and carbohydrate tested (NS).

Discussion

We have validated the aCPQ, a questionnaire developed to assess the presence and severity of abdominal symptoms after ingestion of poorly absorbable carbohydrates. The validation followed previously established procedures [20]. The aCPQ was shown to have excellent psychometric properties and a minimal burden on the patient and

Table 6. Responsiveness of the questionnaire after a carbohydrate challenge

Minutes (mean ± SD)	0	30	60	120	180	360	540
Global Sx score	68.0 ± 72.6	71.6 ± 77.0	69.4 ± 78.0	76.8 ± 88.4	70.6 ± 89.0	91.9 ± 98.5	76.3 ± 89.4
P (global Sx)	**0.006**	0.20	**0.001**	0.004	**0.03**	<0.001	
Cohen's d	0.05	0.03	0.09	0.07	0.24	0.16	
Pain score	15.2 ± 18.8	17.6 ± 21.2	17.3 ± 21.0	17.0 ± 22.4	15.4 ± 22.3	19.0 ± 24.7	15.7 ± 21.8
P	**0.001**	0.19	0.50	0.004	0.59	<0.001	
Cohen's d	0.13	0.02	0.01	0.07	0.16	0.13	
Nausea score	9.3 ± 17.3	12.4 ± 20.0	10.9 ± 19.2	10.4 ± 18.8	8.9 ± 18.1	8.9 ± 17.4	7.7 ± 16.6
P	**0.001**	0.06	0.96	0.001	0.07	0.23	
Cohen's d	0.18	0.08	0.03	0.08	0.00	0.07	
Meteorism score	19.9 ± 23.5	20.2 ± 23.9	20.7 ± 24.3	21.3 ± 25.3	19.7 ± 25.4	23.7 ± 26.5	20.1 ± 25.2
P	0.69	0.50	0.6	0.02	**0.01**	<0.001	
Cohen's d	0.01	0.02	0.02	0.06	0.15	0.13	
Flatulence score	16.9 ± 23.9	14.2 ± 21.8	14.8 ± 22.3	17.1 ± 23.9	16.4 ± 24.2	22.3 ± 28.0	19.3 ± 25.1
P	<0.01	0.46	**0.02**	0.29	**<0.001**	0.05	
Cohen's d	0.11	0.03	0.10	0.03	0.25	0.11	
Diarrhea score	8.9 ± 20.0	7.9 ± 20.3	8.9 ± 21.2	11.8 ± 24.9	9.9 ± 23.2	17.3 ± 29.0	11.4 ± 24.0
P	0.32	0.42	**0.01**	0.03	**<0.001**	<0.001	
Cohen's d	0.05	0.05	0.14	0.08	0.32	0.20	

Exploration of the difference in rating of symptom severity across subsequent tests by 342 patients. Secondary analysis was performed after primary one-way-repeated ANOVA test statistic confirmed one or more significant differences during the observation time.
Global Sx, global symptoms; *P* values (Wilcoxon singed-rank test) and Cohen's d for the comparisons between the two nearest time points (min 0 vs. min 30; min 30 vs. min 60; etc). Significant increases of symptom scores from one time-point to the next are tagged bold.

Table 7. Number of patients with positive or negative hydrogen breath tests (malabsorption) and adult Carbohydrate Perception Questionnaire (intolerance) in the Vienna-original group (a), the Vienna-cross group (b), and the Graz-external group (c)

	Malabsorption		
A, Vienna-original group (*n* = 342)			
Intolerance			
	No	Yes	Sum
No	114	49	163
Yes	82	97	179
Sum	196	146	342
B, Vienna-cross group (*n* = 182)			
	No	Yes	Sum
No	79	26	105
Yes	38	39	77
Sum	117	65	182
C, Graz-external group (*n* = 156)			
	No	Yes	Sum
No	59	18	77
Yes	39	40	79
Sum	98	58	156

resources, as it is brief and easy to administer, fill out, score, and interpret.

The need for a validated instrument has evolved from the clinical importance assigned to the validated diagnosis of carbohydrate intolerance and the evaluation of treatment effects on one hand [12,15,16], and from the lack of available validated symptom questionnaires on the other hand. Intolerance of carbohydrates, among them lactose, has the potential to afflict the majority of world population [21]. Past symptom assessments were mostly non-standardized and may have been subject to doctor- and patient-related biases [7,9,22–24], which limit the confidence on reported results. The only questionnaire related to carbohydrate-induced symptoms that have somehow been validated in the past [25] has been designed to screen patients for a lactose breath test and included vomiting in a summation score, a symptom hardly useful in carbohydrate-induced perception; thus, this scale has not found far-reaching dissemination.

The aCPQ was developed by gastroenterologists who have a longstanding experience in carbohydrate malabsorption and intolerance [1–5,8] and the development of questionnaires [5,26]. The questionnaire was developed after focus group-style interviews with target group representatives, that is, patients with suspected carbohydrate malabsorption or intolerance, and a literature search. Testing of the questionnaire in large patient cohorts included validation of its reliability, different aspects of construct validity, and external validation.

Key properties of a symptom-assessment instrument are its reliability and validity in the population to be studied. Our study population represented patients referred for evaluation of gastrointestinal symptoms suggestive of carbohydrate intolerance after organic disorders had been ruled out. According to our a priori definitions, 43% of patients had malabsorption as demonstrated by a positive breath-H_2 test; 66% of these patients had carbohydrate intolerance and 34% had no symptoms of intolerance. Intolerance, as defined by a positive aCPQ test, was diagnosed in 52% of patients; only 54% of intolerant patients had malabsorption, whereas 46% of intolerant patients had no malabsorption (no H_2-increase ≥20 ppm).

Although the absence of a significant increase in breath hydrogen in patients with intolerance may be partly due to hydrogen-nonexcretion which occurs in up to 20% of patients [27], a significant proportion of intolerant patients remain in whom symptoms cannot be explained by hydrogen-nonexcretion. Hydrogen-nonexcretion is due to activity of methane or sulfide producing colonic bacteria [28]. We have not measured breath-methane and have not quantified breath-CO_2 to adjust for non-alveolar dilution of exhaled air [17] which may have accounted for some 'false negative' hydrogen tests. However, the proportion of patients with isolated elevation of methane is small, and combined measurement of H_2 and methane has shown comparable results with respect to a poor association between malabsorption and clinical symptoms [4,7,16].

We have assessed several types of validity in this study. Construct validity was confirmed by correlation with the results of blinded physicians' interviews, reliability was established by applying the test twice. The time frame was

chosen to be short enough to avoid instability of abdominal symptoms caused by carbohydrates on one hand but at the possible cost of remembering previous answers on the other hand. However, as the patients were not aware of reliability testing and an intervention (lactose ingestion) separated the two tests, we are confident that patients did not intentionally duplicate the questionnaires. Cronbach's alpha, a commonly used measure of internal consistency, equaled 0.85, which is generally regarded as good [29]. Although Cronbach's alpha is not necessarily relevant, as the aCPQ is only reported as single items, future studies may show that a total or average score of gastrointestinal symptoms severity might be of clinical relevance.

The aCPQ allows for the determination of both the quality and severity of abdominal symptoms after a carbohydrate challenge. A VAS was used for the scaling of symptom severity. The documented benefits of VAS scaling in gastrointestinal and extra-gastrointestinal diseases highly outweigh the potential problems [30–32]. Several established questionnaires measure gastrointestinal symptoms using a VAS approach, such as the symptom severity scale (Francis *et al* 1997) or the VAS-IBS (Bengtsson *et al*. 2011). While these scales assess symptoms on a medium-term timeframe (ten to 14 days), the aCPQ needs responsiveness, which is the ability of the questionnaire to detect changes over time, in the short-term. Responsiveness was demonstrated for each symptom in patients who retrospectively reported the manifestation of symptoms during the 3 h of the breath test.

The specific symptom that led to the diagnosis of intolerance differed widely among subjects and often more than one symptom increased by ≥20 mm over baseline during the course of the breath test, with flatulence (50%) and pain (47%) being the most frequent individual symptoms leading to the diagnosis of intolerance.

The combination of breath tests with an unbiased symptom assessment identifies four different entities after a carbohydrate load: (1) malabsorption plus symptoms, (2) malabsorption only, (3) symptoms only, and (4) none of the above. While in the past carbohydrate challenges and breath tests have focused on malabsorption, recent data suggest that it is not malabsorption but symptom development which is of superior clinical relevance [5,10] including starting treatment [15].

It has been suggested that the term 'carbohydrate intolerance' should be replaced by 'sensitivity to the carbohydrate' or 'carbohydrate hypersensitivity' [5], as there is confusion regarding the term 'carbohydrate intolerance': the term intolerance has often been used indiscriminately in the context of carbohydrate malabsorption, encompassing both malabsorption with or without documented symptoms. Data accumulate that suggest that carbohydrate-induced symptoms are mainly due to visceral hypersensitivity [5,7]; thus, 'carbohydrate hypersensitivity' expresses the link to visceral hypersensitivity, which is an established term in functional gastrointestinal research [33].

In summary, we have developed a novel instrument, the aCPQ, which allows for standardized, unbiased assessment of symptom severity during carbohydrate breath tests and therefore allows a valid diagnosis of carbohydrate intolerance. The questionnaire is available in German and will have to undergo a standard translation process before it is valid to use in other languages and cultures [34]. It may set an imperatively needed standard for the diagnosis of carbohydrate intolerance and is suitable for clinical testing and therapeutic trials and research.

Acknowledgements

J.H. involved in study concept and design, study supervision, statistical analysis and interpretation of data, drafting of the manuscript, and critical revision of the manuscript for important intellectual content. M.S. participated in acquisition of data, administrative, technical support, and revision of the manuscript. K.M.E., G.R., and S.L. contributed to acquisition of data. H.F.H. involved in study design, acquisition of data, administrative, technical, or material support, and revision of the manuscript. Provision of questionnaires: For studies without financial support from industrial sponsors, the questionnaires are provided free of charge.

Conflicts of interest

J.H. and H.F.H. are shareholders: Carboception. For the remaining authors, there are no conflicts of interest.

References

1 Hammer HF. Colonic hydrogen absorption: quantification of its effect on hydrogen accumulation caused by bacterial fermentation of carbohydrates. *Gut* 1993; 34:818–822.
2 Hammer HF, Fine KD, Santa Ana CA, Porter JL, Schiller LR, Fordtran JS. Carbohydrate malabsorption. Its measurement and its contribution to diarrhea. *J Clin Invest* 1990; 86:1936–1944.
3 Hammer HF, Santa Ana CA, Schiller LR, Fordtran JS. Studies of osmotic diarrhea induced in normal subjects by ingestion of polyethylene glycol and lactulose. *J Clin Invest* 1989; 84:1056–1062.
4 Hammer HF, Petritsch W, Pristautz H, Krejs GJ. Assessment of the influence of hydrogen nonexcretion on the usefulness of the hydrogen breath test and lactose tolerance test. *Wien Klin Wochenschr* 1996; 108:137–141.
5 Hammer V, Hammer K, Memaran N, Huber WD, Hammer K, Hammer J. Relationship between abdominal symptoms and fructose ingestion in children with chronic abdominal pain. *Dig Dis Sci* 2018; 63:1270–1279.
6 Major G, Pritchard S, Murray K, Alappadan JP, Hoad CL, Marciani L, *et al*. Colon hypersensitivity to distension, rather than excessive gas production, produces carbohydrate-related symptoms in individuals with irritable bowel syndrome. *Gastroenterology* 2017; 152:124.e2–133.e2.
7 Wilder-Smith CH, Materna A, Wermelinger C, Schuler J. Fructose and lactose intolerance and malabsorption testing: the relationship with symptoms in functional gastrointestinal disorders. *Aliment Pharmacol Ther* 2013; 37:1074–1083.
8 Hammer HF, Hammer J. Diarrhea caused by carbohydrate malabsorption. *Gastroenterol Clin North Am* 2012; 41:611–627.
9 Wilder-Smith CH, Olesen SS, Materna A, Drewes AM. Repeatability and effect of blinding of fructose breath tests in patients with functional gastrointestinal disorders. *Neurogastroenterol Motil* 2019; 31:e13497.
10 Wilder-Smith CH, Olesen SS, Materna A, Drewes AM. Fermentable sugar ingestion, gas production, and gastrointestinal and central nervous system symptoms in patients with functional disorders. *Gastroenterology* 2018; 155:1034.e6–1044.e6.
11 Tuck CJ, McNamara LS, Gibson PR. Editorial: rethinking predictors of response to the low FODMAP diet – should we retire fructose and lactose breath-hydrogen testing and concentrate on visceral hypersensitivity? *Aliment Pharmacol Ther* 2017; 45:1281–1282.
12 Suchy FJ, Brannon PM, Carpenter TO, Fernandez JR, Gilsanz V, Gould JB, *et al*. National Institutes of Health Consensus Development Conference: lactose intolerance and health. *Ann Intern Med* 2010; 152:792–796.

13 Deng Y, Misselwitz B, Dai N, Fox M. Lactose intolerance in adults: biological mechanism and dietary management. *Nutrients* 2015; 7:8020–8035.
14 Chumpitazi BP, Shulman RJ. Dietary carbohydrates and childhood functional abdominal pain. *Ann Nutr Metab* 2016; 68 (Suppl 1):8–17.
15 Gasbarrini A, Corazza GR, Gasbarrini G, Montalto M, Di Stefano M, Basilisco G, *et al*.; 1st Rome H2-Breath Testing Consensus Conference Working Group. Methodology and indications of H2-breath testing in gastrointestinal diseases: the Rome Consensus Conference. *Aliment Pharmacol Ther* 2009; 29 (Suppl 1):1–49.
16 Rezaie A, Buresi M, Lembo A, Lin H, McCallum R, Rao S, *et al*. Hydrogen and methane-based breath testing in gastrointestinal disorders: the North American Consensus. *Am J Gastroenterol* 2017; 112:775–784.
17 Hammer J, Hammer HF. There is an unmet need for test-specific, validated symptom questionnaires for breath tests in adults. *Gastroenterology* 2019; 156:1220–1221.
18 Lawshe CH. A quantitative approach to content validity. *Pers Psychol* 1975; 28:563–575
19 Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* 1994; 6:284–290.
20 Artino AR Jr, La Rochelle JS, Dezee KJ, Gehlbach H. Developing questionnaires for educational research: AMEE Guide No. 87. *Med Teach* 2014; 36:463–474.
21 Johnson JD. The regional and ethnic distribution of lactose malabsorption. In: Paige DM, Bayless TM, editors. *Lactose digestion clinical and nutritional implications*. Baltimore, Maryland: Johns Hopkins University Press; 1981. pp. 11–22.
22 Beyerlein L, Pohl D, Delco F, Stutz B, Fried M, Tutuian R. Correlation between symptoms developed after the oral ingestion of 50 g lactose and results of hydrogen breath testing for lactose intolerance. *Aliment Pharmacol Ther* 2008; 27:659–665.
23 Yang J, Fox M, Cong Y, Chu H, Zheng X, Long Y, *et al*. Lactose intolerance in irritable bowel syndrome patients with diarrhoea: the roles of anxiety, activation of the innate mucosal immune system and visceral sensitivity. *Aliment Pharmacol Ther* 2014; 39:302–311.
24 Tomba C, Baldassarri A, Coletta M, Cesana BM, Basilisco G. Is the subjective perception of lactose intolerance influenced by the psychological profile? *Aliment Pharmacol Ther* 2012; 36: 660–669.
25 Casellas F, Varela E, Aparici A, Casaus M, Rodríguez P. Development, validation, and applicability of a symptoms questionnaire for lactose malabsorption screening. *Dig Dis Sci* 2009; 54:1059–1065.
26 Koloski NA, Jones M, Hammer J, von Wulffen M, Shah A, Hoelz H, *et al*. The validity of a new Structured Assessment of Gastrointestinal Symptoms Scale (SAGIS) for evaluating symptoms in the clinical setting. *Dig Dis Sci* 2017; 62:1913–1922.
27 Gilat T, Ben Hur H, Gelman-Malachi E, Terdiman R, Peled Y. Alterations of the colonic flora and their effect on the hydrogen breath test. *Gut* 1978; 19:602–605.
28 Gibson GR, Cummings JH, Macfarlane GT, Allison C, Segal I, Vorster HH, Walker AR. Alternative pathways for hydrogen disposal during fermentation in the human colon. *Gut* 1990; 31:679–683.
29 Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J Pers Assess* 2003; 80:99–103.
30 McCormack HM, Horne DJ, Sheather S. Clinical applications of visual analogue scales: a critical review. *Psychol Med* 1988; 18:1007–1019.
31 Gorrall BK., Curtis JD, Little TD, Panko P. Innovations in measurement: visual analog scales and retrospective pretest self-report designs. *Actualidades en Psicología* 2016; 30:2–7
32 Couper MP, Tourangeau R, Conrad FG, Singer E. Evaluating the effectiveness of visual analog scales: a web experiment. *Soc Sci Comput Rev* 2006; 24: 227–245.
33 Delgado-Aros S, Camilleri M. Visceral hypersensitivity. *J Clin Gastroenterol* 2005; 39:S194–S203; discussion S210.
34 Sperber AD. Translation and validation of study instruments for cross-cultural research. *Gastroenterology* 2004; 126:S124–S128.