

Search history 2017	
<p>PREVALENCE (Q1): 10159 references imported for screening ↓ 5 duplicates removed 10154 studies screened against title and abstract ↓ 10062 studies excluded 92 studies assessed for full-text eligibility ↓ 81 studies excluded 35 not about Prevalence 29 adult population 13 poor study design 1 same publication was listed twice 1 wrong outcomes 1 wrong patient population 1 wrong study design ↓ 0 studies ongoing 0 studies awaiting classification 11 studies included 2 studies added by manual search</p>	<p>TESTS (Q2-6): Q2: 2475 references imported for screening ↓ 1 duplicate removed 2474 studies screened against title and abstract ↓ 2368 studies excluded 106 papers for all diagnostic questions (Q2-Q6). ↓ Abstract review 75 excluded; did not answer Q2 31 papers remain Review of full manuscript 25 papers excluded: 6 remaining studies 2 recently published studies (2017) included as relevant for the questions ↓ 8 studies / original publications were considered</p>
<p>Q3: 2475 references imported for screening ↓ 1 duplicate removed 2474 studies screened against title and abstract ↓ 2368 studies excluded 106 studies review of abstract ↓ 87 studies excluded 19 studies assessed by full text for eligibility ↓ 15 studies excluded: no detailed data on TGA titer 6 studies added published after initial search 1 study added based on reference in included study 0 studies ongoing 0 studies awaiting classification 11 studies included</p>	<p>Q4: 4022 in search result ↓ 1548 duplicates removed 2474 studies screened against title and abstract ↓ 2474 studies excluded 0 studies assessed for full-text eligibility ↓ 0 studies excluded 0 studies ongoing 0 studies awaiting classification 0 studies included 18 studies re-evaluated from manual search</p>
<p>Q5: 2475 references imported for screening plus 5 studies identified by manual search ↓ 1 duplicate removed 2479 studies screened against title and abstract ↓ 2372 studies excluded 107 studies assessed by full-text for eligibility ↓ 10 studies included</p>	<p>Q6: 2475 references imported for screening plus 5 studies identified by manual search ↓ 1 duplicate removed 2479 studies screened against title and abstract ↓ 2369 studies excluded 110 studies assessed by full-text for eligibility ↓ 88 studies excluded (60 in Covidence, 28 during extraction of data) 66 not answering Q5 4 adult population 8 Mixed (adult + paediatric) population where data for children were not separately extractable 2 No detailed antibody results, number of true and false positives not extractable 1 Not commercially available test 4 TGA-IgA test not suitable for the calculation of xULN 2 TGA-IgA test result part of the reference standard 1 Reference standard not applied to all participants ↓ 22 studies included</p>
<p>BIOPSY (Q7-10) Q7: 1097 references imported for screening ↓ 5 duplicates removed 1092 studies screened against title and abstract ↓ 1067 studies excluded 25 studies assessed for full-text eligibility ↓ 23 studies excluded 20 not relevant for the questions. 1 adult population 1 no readable abstract available 1 duplicate ↓ 0 studies ongoing 0 studies awaiting classification 8 additional studies identified, 2 of which from 2017 10 studies included</p>	<p>Q8: 1097 references imported for screening ↓ 5 duplicates removed 1092 studies screened against title and abstract ↓ 1067 studies excluded 25 studies assessed for full-text eligibility ↓ 4 studies excluded as not relevant for Q8 6 studies excluded as adult studies 15 studies read as full-text and evaluated on Quadas 2 analysis (2 studies on inter-observer agreement and 13 on bulb histopathology) 1 more study added later as histology interobserver study in children (Villanaci, et al 2018) –this study was not included in the initial analysis</p>
<p>Q9: 1097 references imported for screening ↓ 12 studies assessed for full-text eligibility ↓ 6 studies excluded as performed in adult population 6 studies included</p>	<p>Q10: 1097 references imported for screening ↓ 5 duplicates removed 1092 studies screened against title and abstract ↓ 1067 studies excluded 25 studies assessed for full-text eligibility ↓ 18 studies excluded 16 not relevant for the question 1 adult population 1 duplicate ↓ 0 studies ongoing 0 studies awaiting classification 6 studies included</p>

Figure S1: Literature Search Results

Figure S2: Summary ROC curves showing summary points with corresponding 95% confidence intervals for TGA-IgA, DGP-IgG and EMA-IgA

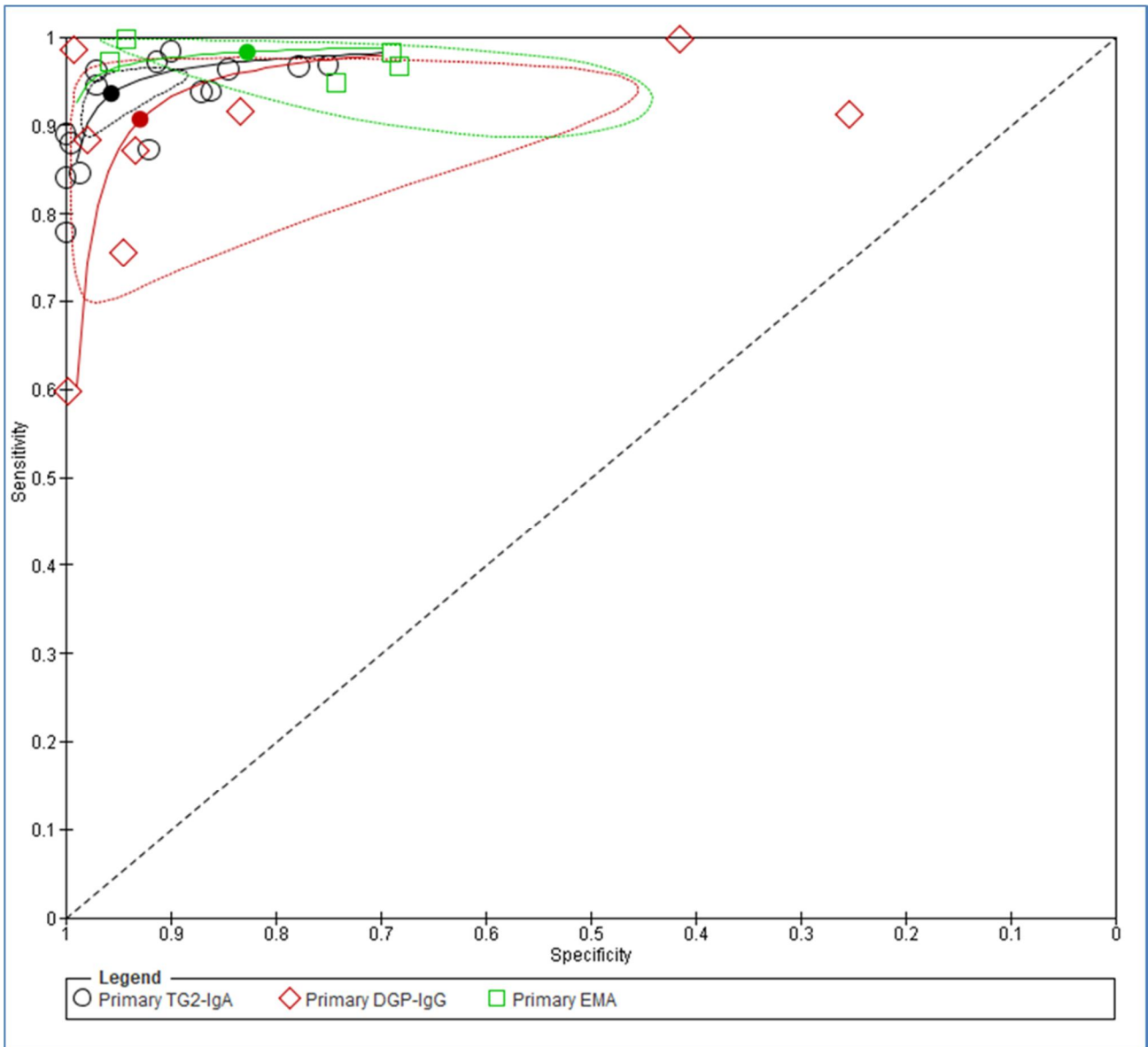


Table S1: QUADAS2 analysis of Question 1

Study	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Agardh 2015	Low	Low	Low	Low	High	Low	Unclear
Bramanti 2014	Low	Low	Low	Low	Low	Low	Low
Cristofori 2014	Unclear	Low	Low	Low	Low	Low	Low
Dehghani 2015	High	Low	Low	Low	Low	Low	Unclear
Fitzpatrick 2001	High	High	High	Low	High	High	High
Imanzadeh 2005	Low	Low	Low	Low	Low	Low	Low
Kalayci 2005	High	Low	Unclear	Low	Unclear	Low	Unclear
Kansu 2015	Low	Low	Low	Low	Low	Low	Unclear
Khatib 2016	Unclear	Unclear	Low	Low	Unclear	Low	Low
Lass 2015	Unclear	Low	High	Low	High	Low	Unclear
Sattar 2011	High	Unclear	Unclear	Unclear	Low	Unclear	Unclear
Shakeri 2009	High	Low	Unclear	Low	Low	Low	Unclear
Sharma 2007	Unclear	Unclear	Unclear	Unclear	Low	Low	Unclear

Table S3: QUADAS2 analysis of Question 3

Study	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
<i>Retrospective</i>							
Donat 2012	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear
Nevorat 2013	Unclear	Low	Low	Unclear	Low	Low	Low
Trovato 2015	Unclear	Low	Unclear	Unclear	Low	Low	Low
<i>Mass screening</i>							
Webb 2015	Low	Low	Low	Low	Low	Low	Low
Jansen 2017	Low	Low	Unclear	Unclear	Low	Low	Low
<i>Prospective</i>							
Lionetti 2014	Low	Low	Low	Low	Low	Low	Low
Vriezinger 2014	Low	Low	Low	Low	Low	Low	Low
Cilleruelo 2016	Low	Low	Low	Low	Low	Unclear	Low
Werkstetter 2017	Low	Low	Low	Low	Low	Low	Low
Wolf 2017	Low	Low	Low	Low	Low	Low	Low
Paul 2017	Unclear	Low	Unclear	Low	Unclear	Low	Low

Table S5: QUADAS2 analysis of Question 5

Study	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Prospective							
Horwitz 2015	Low	Low	High	High	High Adults	Low	Low
Vriezinga 2014	Low	Low	Low	Low	Low	Low	Low
Wolf 2017	Low	Low	Low	Low	Low	Low	Low
Retrospective							
Aberg 2009	Low	Low DGP	High	High	Low	Low	Low
Absah 2017	Low	Low TGA-IgG in TGA-IgA neg	High	High	High Children & adults	Low	Low
Foucher 2012	Low	Low AGA-IgA	High	High	High < 2 yrs.	Low	High
Frulio 2015	Low	High	Unclear	Unclear	High <4 (A) & <2yrs.(B)	High Cut off defined by ROC curve	Unclear
Hojsak 2012	Unclear	Low TGA-IgG&IgA EMA- IgG&IgA	High	High	High >3 yrs	High	High
Parizade 2010	Low	Low	High	High	High <2 yrs.	High	Low

Vermeersch	Low/Unclear	Low DGP-IgG TGA/DGP combination	Unclear	Unclear	High Children and adults	Low	High
-------------------	-------------	------------------------------------------	---------	---------	--------------------------------	-----	------

Table S7: QUADAS2 analysis of Question 7

Study	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Prospective studies							
Mubarak 2013	Low	Low	Low	Low	Low	Low	Low
Wolf 2017	Low	Low	Low	Low	Low	Low	Low
Werkstetter 2017	Low	Low	Low	Low	Low	Low	Low
Retrospective studies							
Donaldson 2008	Unclear	Low	Low	High	Unclear	High	Low
Mubarak 2011	Low	Low	Low	Low	Low	Low	Low
Panetta 2011	Low	Unclear	Low	Unclear	Low	Low	Low
Bürgin-Wolff 2013	High	Low	Low	Unclear	Low	Low	Low
Klapp 2013	Low	Low	Unclear	Low	Low	Low	Unclear
Gidrewicz 2015	Low	Low	Unclear	Unclear	Low	Low	Unclear
Nevoral 2013	Low	Low	Unclear	Low	Low	Low	Unclear

Table S9: QUADAS2 analysis of Question 9

	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Tosco 2008	Unclear	Low	Unclear	Low	Low	Low	Unclear
Tosco 2013	Unclear	Low	High	Low	Low	High	Unclear
Tosco 2015	Unclear	Unclear	Unclear	Low	Unclear	Unclear	Unclear
Koskinen 2008	Unclear	Low	Unclear	Low	Unclear	Unclear	Unclear
Borrelli 2010	Unclear	Low	High	Low	Low	Unclear	Unclear
Maglio 2011	Low	Low	Unclear	Low	Low	Low	Unclear

Table S11: GRADE analysis of Question no. 1

Evaluation by: Serious, not serious, none

GRADE analysis of diagnostic tests (BMJ 2008;336:1106)

No. of studies	Study design	Indirectness		Inconsistency	Imprecision	Publication bias	Total no. of patients	Quality (1-4 high or low)
		Outcomes	Patient population, comparisons					
13	10 prospective	10 Yes: Diagnosis	Variable (CD patients, potential CD, at risk patients, controls)	Not serious	Not serious	Not serious	7198 CD 537	Low-moderate 3
	3 retrospective	3 Yes: Diagnosis	Variable	Not serious	Not serious	Not serious	13,079 CD 256	Moderate 3

Table S12: GRADE analysis of Question no. 2

Evaluation by: Serious, not serious, none

GRADE analysis of diagnostic tests (BMJ 2008;336:1106)

No. of studies	Study design	Indirectness		Inconsistency	Imprecision	Publication bias	Total no. of CD patients / controls	Quality (1-4 high or low)
		Outcomes	Patient population, comparisons					
8	5 prospective	None	None	None	None	Unknown	5170 / unknown	High (4)
	3 retrospective	Not serious/serious	Not serious/serious	Not serious	Not serious	Unknown		Low-moderate (2)

	Biopsied	HLA-typed	HLA+ and CD	HLA-neg CD
Werkstetter	707	707	645	0
Wolf et al	898	449	277	0
Clouzeau-Girard	162	162	81	0
Kurppa et al.	140	140	114	0
Sandstrom et al.	184	184	153	0
Donat - 1st possibility	2177	751*	401	9
Donat - 2nd possibility	2177	1467	1467	28
				0 to 3 (final diagnosis unclear)
Klapp et al.	150	150	133	
Tucci et al.	749	368	310	7
TOTALS	5170			

* plus TGA-IgA and EMA-IgA performed

Table S13: GRADE analysis of Question no. 3

Evaluation by: Serious, not serious, none

GRADE analysis of diagnostic tests (BMJ 2008;336:1106)

No. of studies	Study design	Indirectness		Inconsistency	Imprecision	Publication bias	Total no. of patients /controls	Quality (1-4 high or low)
		Outcomes	Patient population, comparisons					
11	2 cross-sectional (mass screening)	Provided	Good	Largely consistent findings	None	Unknown	CD: 555	High 4
	6 prospective	Provided	Variable (at risk population, general population, suspected CD)	Largely consistent findings	None	Unknown	Controls: None	
	3 retrospective	Provided	Variable (suspected CD, multicenter/single centre)	Variable for retrospective	Not serious	Unknown		

Table S14: GRADE analysis of Question no. 4

Evaluation by: Serious, not serious, none

GRADE analysis of diagnostic tests (BMJ 2008;336:1106)

No. of studies	Study design	Indirectness		Inconsistency	Imprecision	Publication bias	Risk of bias	Total no. of CD patients / controls	Quality (1-4 high or low)
		Outcomes	Patient population, comparisons						
18	Cohort or cross-sectional	Sens Spec	Children referred due to suspicion of CD, largely comparable	No	Narrow confidence limits anticipated	No	Low	3332/3759	3 high

differs across studies

** reference standard is in all cases the histological analysis of duodenal biopsy. Two papers draw the attention of the error rate of the biopsy, 4-5% (Wolf 2017; Werkstetter 2017)

Table S15: GRADE analysis of Question no. 5

Evaluation by: Serious, not serious, none

GRADE analysis of diagnostic tests (BMJ 2008;336:1106)

No. of studies	Study design	Indirectness		Inconsistency	Imprecision	Publication bias	Total no. of CD patients / controls	Quality (1-4 high or low)
		Outcomes	Patient population, comparisons					
10	3 prospective	3 with necessary outcomes	1 of 3. 1 adult unbiased, 2 paediatric, 1 unbiased	Consistent findings	Not serious	Unknown	466 / 3846	Moderate to high (3)
	7 retrospective	6 with necessary outcomes	7 of 7, all paediatric but all with high selection bias	Largely consistent findings	Variable	Unknown	Cannot be calculated	Low-moderate (2)

Table S16: GRADE analysis of Question no. 6

Evaluation by: Serious, not serious, none

GRADE analysis of diagnostic tests (BMJ 2008;336:1106)

No. of studies	Study design	Indirectness		Inconsistency	Imprecision	Publication bias	Total no. of CD patients/controls	Quality (1-4 high or low)
		Outcomes	Patient population, comparisons					
19 [36 datasets]	Retrospective case-control	Diagnosis yes/no	Relevant	Serious	Unknown	Unkown	3636/2370	Low or moderate (large effect, large sample size) (2)
3 [11 datasets]	Prospective cohorts	Diagnosis yes/no	Relevant	Not serious	Not serious	Not serious	1235/440	High (4)

Table S17: GRADE analysis of Question no. 7

Evaluation by: Serious, not serious, none

GRADE analysis of diagnostic tests (BMJ 2008;336:1106)

No. of studies	Study design	Indirectness		Inconsistency	Imprecision	Publication bias	Total no. of CD patients/controls	Quality (1-4 high or low)
		Outcomes	Patient population, comparisons					
10	3 cross-sectional/prospective 7 retrospective	Yes: diagnosis yes/no	No, for the 3 prospective studies Variable for the 7 retrospective	Largely consistent findings	Not assessed, no meta-analysis	Unknown	1357/457 Prospective only	High (4)

Table S18: GRADE analysis of Question no. 8

Evaluation by: Serious, not serious, none

GRADE analysis of diagnostic tests (BMJ 2008;336:1106)

No. of studies	Study design	Indirectness		Inconsistency	Imprecision	Publication bias	Total no. of CD patients / controls	Quality (1-4 high or low)
		Outcomes	Patient population, comparisons					
13 on bulb Histopathology	9 prospective 4 retrospective	The correct diagnosis	For 8 studies representative samples were taken For the remaining studies the samples vary in selection and may not be representative for the study question	The findings are partially controversial, formal meta-analysis has not been done	Imprecision has not been formally assessed	None detected	2708 /439	2 studies high quality (4), 6 studies good quality (3) 5 studies lower quality (2)
3 on interobserver agreement	Prospective	The correct diagnosis	Representative samples were taken	Partially controversial findings	Imprecision has not been formally assessed	None detected		High quality studies(4)

Table S19: GRADE analysis of Question no. 9

Evaluation by: Serious, not serious, none

GRADE analysis of diagnostic tests (BMJ 2008;336:1106)

No. of studies	Study design	Indirectness		Inconsistency	Imprecision	Publication bias	Total no. of patients / controls	Quality (1-4 high or low)
		Outcomes	Patient population, comparisons					
6	2 cross-sectional/prospective 4 retrospective	Yes: diagnosis yes/no	Yes for the prospective studies Variable for retrospective	Largely consistent findings	Not assessed, No meta-analysis	Unknown	CD (incl potential CD): 465 Controls: 271	Moderate to high (1) Low-moderate (5)

Table S20: GRADE analysis of Question no. 10

Evaluation by: Serious, not serious, none

GRADE analysis of diagnostic tests (BMJ 2008;336:1106)

No. of studies	Study design	Indirectness		Inconsistency	Imprecision	Publication bias	Total no. of CD patients /controls	Quality (1-4 high or low)
		Outcomes	Patient population, comparisons					
6	1 prospective 1 cross-sectional 4 retrospective	High risk of bias Findings by endoscopy other than coeliac disease Not evaluated – outcomes not reported of additional findings	1 predominantly adult population 3 including only paediatric cases 2 including children and young adults	Largely consistent findings	Not assessed No meta-analysis	Unknown	2383 / 90113	Low (1)

Table S21: Outcome of small bowel biopsies in asymptomatic children with TGA-IgA $\geq 10 \times$ ULN

Study	Methods	No of patients	No of biopsies	Marsh 0-1	Marsh 2-3	n=
Retrospective						
Nevoral 2013	Retrospective Suspected CD Single center, Czech Republic	114	114	11	103	114
Trovato 2015	Retrospective Diagnosed CD Single center, Italy	40	40	3	37	40
Donat 2016	Retrospective Consecutive cases of suspected CD Multicenter, Spain	69	69	4	65	69
Cross-sectional						
Webb 2015	Mass screening 2 Separate cohorts of 12 year olds, Sweden	64	64	1	63	64
Jansen 2017	Mass screening Birth cohort, 6 and 9 year olds, The Netherlands	20	19	3	16	20
Prospective						
Lionetti 2014	Birth cohort first degree relative with CD Multicenter, Italy	24	24	3	21	24
Vriezinga 2014	Birth cohort first degree relative with CD Multicenter, International	29	27	0	27	29
Cilleruelo 2016	Birth cohort HLA- DQ2 and/or DQ8 +,	13	13	4	9	13

	2-3 year olds Single center, Spain					
Werkstetter 2017	Consecutive Suspected CD Multicenter, International	51	51	1*	50*	51
Wolf 2017	Consecutive Suspected CD Multicenter, International	47	47	2	45	47
Paul 2017	Consecutive Suspected CD Single center, United Kingdom	84	84	0	84	84
Total		555	552	32	520	555

*In this study biopsies were blindly evaluated by two pathologists. Discrepant Marsh classification (Marsh 0-1 versus Marsh 2-3) were found in 7.1%. Therefore, the final diagnosis of each case considered not only histology, but also TGA and EMA testing. In 50 of 51 asymptomatic children with TGA ≥ 10 ULN the final diagnosis was CD, while in one child the diagnosis remained inconclusive.

Supplementary Material S22: Additional information on retrospective studies summarized in the manuscript regarding question 5.

The number of the references are related to the main manuscript:

Aberg et al (48) performed serological testing in all children below 3 years of age with available stored serum samples (1382/1661) for DGP IgG/IgA and TGA-IgA/DGP IgG/IgA. Patients with IgA deficiencies were excluded. Of 167 children with a positive result in any of the tests, only 32 underwent biopsies. None of the children with biopsy confirmed CD was positive of DGP IgG/IgA and negative for TGA-IgA. The results indicate that the screening for young children (below 4 years of age) should be performed with TGA-IgA, but not DGP based tests.

Friolo et al (47) reported the results of 730 children between 6 months and 4 years of age (group A, 78 of 730 with biopsy proven CD), thereof 348 were below 2 years of age (group B, 21 of 348 with biopsy proven CD), who have been tested for TGA-IgA and DGP-IgA and DGP-IgG in their laboratory within a 2 year period. A drawback of this study was that for each test the optimal cut off was defined by ROC curves for the two 2 age groups, which was for all three tests higher in the older (group A) compared to the younger cohort (group B). In both age groups the sensitivity and specificity of TGA-IgA was higher compared to the other two tests. The results indicate that the screening for young children (below 4 years of age) should be performed with TGA-IgA, but not with DGP based tests.

Hojdak et al (46) analyzed the serological data of children below 3 years of age tested in Israel during a defined time period. Of 6074 included patients 4085 were also tested for DGP antibodies, with 232 of them having positive results. Unfortunately a large limitation of the study was that only 59/232 children with positive results for EMA, TGA or DGP underwent biopsies. Histopathology indicated CD in 47/59 cases (31/47 had all 3 tests performed) and no CD in the remaining 12 (9/12 with all 3 tests performed). Again, neither reference histology nor a challenge and re-biopsy procedure in seronegative children had been performed. In addition, total IgA was known in only 50 of the 59 children used in the final analysis. With the cut off given by the manufacturer sensitivity was high for EMA IgA & IgG (96%), TGA-IgA (97%) and DGP IgA & IgG (100%) while specificity showed marked differences (91%, 50% and 44%, respectively). In those 9 patients with normal histology and results of all three tests available, one child was positive for EMA, 3 for TGA and 7 for DGP antibodies.

Parizade et al. (42) tested all serum samples from children below 2 years of age (n= 5036) which were sent to the laboratory over a period of 17 months for CD serology for TGA-IgA and total IgA, and for DGP-IgA and IgG. Of 202 children with a positive DGP results, 35 were also positive for TGA, and 16 with negative TGA-IgA result were IgA deficient. Of the remaining 152 children with positive DGP but negative TGA result only 12 underwent biopsies and in 6 patients either histology confirmed or excluded the diagnosis of CD. Serological follow up on a gluten containing diet was available in 68 TGA negative children: DGP decreased or became negative in 49, increased in 13 and in the further 6 children results were not known. Of 152 cases with initial DGP-positive/TGA-negative results only one infant converted to TGA-positive, but turned DGP negative and was confirmed to have CD by biopsies. The authors conclude that in infants <2 years DGP positivity in the absence of TGA is very frequent and in most children a transient phenomenon not predicting CD.

Vermeersch (49) et al analyzed the serological data of 107 CD cases and 542 controls including adults and children. All patients underwent biopsies for histology. Sera of all patients were tested for total IgA, TGA – IgA and DGP-IgG of two different manufacturers and a screening test combining TGA-IgG & IgA with DGP IgA & IgG. For each tests and the combinations of each test the calculated the likelihood ratio (LR) for each test and different combinations. They confirmed that the highest positive LR was reached when TGA-IgA and DGP-IgG were positive and the lowest when both tests were negative. Accordingly for both manufacturers sensitivity increased with applying DGP-IgG in addition to TGA-IgA testing but specificity decreased. For a given pre-test probability the post-test probability of the given combinations depended on the manufacturer. The LR for a positive DGP-IgG result in the absence of TGA-IgA positivity after exclusion of IgA deficient cases was low (5.1 and 1.6 for the two different manufacturers) A strong weakness of this study was that pediatric coeliac patients were later added to the cohort. These patients were not consecutive patients as stated in the manuscript. In addition patients with Marsh 1 and 2 lesions were considered to be CD based on symptoms, other diagnoses and serological response to gluten free diet.

Supplementary Material S23: Statements from the 2012 guidelines still in force

This evidence search only explored 10 selected fields, which does not influence the validity of the following statements which rest on previous evidence found still satisfactory for current times.

3.4.3

(↑↑) Laboratories providing CD antibody test results for diagnostic use should continuously participate in quality control programme at national or European level.

3.4.4

(↑↑) TGA and DGP antibody laboratory test results should be reported as numeric values together with specification of the immunoglobulin class measured, the manufacturer, the cutoff value defined for the specific test kit, and, (if available) the level of 'high' antibody values. It is not sufficient to state only positivity or negativity. Information on the source of the antigen (natural, recombinant, human, non-human) should be provided for in-house methods.

3.4.5

(↑↑) Reports on EMA results should contain the specification of the investigated immunoglobulin class, the interpretation of the result (positive or negative), the cutoff dilution and the specification of the substrate tissue. It is also useful to have the information on the highest dilution still positive.

3.4.12

(↑) The use of tests for the detection of antibodies of any type (IgG, IgA, secretory IgA) in fecal samples are not recommended for clinical evaluation.

3.4.18

(↑↑) Skin immunofluorescent study-proven dermatitis herpetiformis can also be regarded as confirmation of gluten sensitivity (added: independent of serum antibody results).

4.3.4

(↑↑) In seronegative cases with strong clinical suspicion of CD, small intestinal biopsies are recommended.

4.3.10

(↑) Patients fulfilling the diagnostic criteria of CD do not need biopsies on a gluten-free diet (GFD).

4.3.11

(↑) If there is no clinical response to a GFD in symptomatic patients, after a careful dietary assessment to exclude lack of compliance, further investigations are recommended. They may include further biopsies.

4.3.12

(↑) Gluten challenge is not considered mandatory, except under unusual circumstances. These include situations where there is doubt about the initial diagnosis including patients with no CD specific antibodies before starting a GFD.

4.3.13

(↑) If gluten challenge is indicated it should not be recommended before the age of 5-6 years and during the pubertal growth spurt.

4.3.14

(↑↑) It is recommended that gluten challenge is performed under medical supervision preferably by a pediatric gastroenterologist.

4.3.16

(↑↑) The daily dietary intake during gluten challenge is recommended to contain a normal amount of gluten (around 15g/day).

4.3.17

(↑↑) It is recommended that during the challenge period TGA-IgA antibody (IgG in the case of IgA deficiency) is measured. A patient should be considered relapsed (and hence the diagnosis of CD confirmed) if CD serology becomes positive and a clinical and/or histological relapse observed. In the

absence of positive antibodies/symptoms the challenge should be considered over after two years and biopsies performed. Follow up should be continued since relapse may occur after more than two years.