

Accuracy of Diagnostic Antibody Tests for Coeliac Disease in Children: Summary of an Evidence Report

*Klaus Giersiepen, †Monika Lelgemann, *Nina Stuhldreher, ‡Luca Ronfani,
 ||Steffen Husby, §Sibylle Koletzko, and ¶#Ilma R. Korponay-Szabó,
 and the ESPGHAN Working Group on Coeliac Disease Diagnosis

ABSTRACT

Objective: The aim of this study was to summarise the evidence from 2004 to September 2009 on the performance of laboratory-based serological and point of care (POC) tests for diagnosing coeliac disease (CD) in children using histology as reference standard.

Patients and Methods: We searched MEDLINE and EMBASE for studies reporting on children for tests based on IgA and IgG anti-gliadin (AGA), endomysial (EmA), anti-transglutaminase-2 (TG2), and anti-deamidated gliadin peptides (DGP) antibodies or POC tests. For inclusion, histological analysis of duodenal biopsies and sensitivity and specificity for index tests had to be reported. Data were pooled and summary measures calculated for sensitivity, specificity, positive and negative likelihood ratios ("LR+", "LR-"), and diagnostic odds ratios (DOR). In case of elevated statistical heterogeneity, studies reaching 90% sensitivity or specificity were reported.

Results: A total of 2510 articles were reviewed; 16 entered meta-analysis, reporting on 3110 patients (1876 with CD, 1234 without CD). For IgA-EmA, sensitivity was $\geq 90\%$ in 7/11 studies and pooled specificity 98.2%. For IgA-anti-TG2, 11/15 studies yielded sensitivities $\geq 90\%$ and 13/15 specificities $\geq 90\%$. For IgA-DGP, sensitivity ranged between 80.7% and 95.1% (specificity 86.3%–93.1%); for IgG-DGP between 80.1% and 98.6% (specificity 86.0–96.9%). IgA-EmA had the highest pooled DOR (554) and LR+ (31.8) for a laboratory test, followed by IgA-anti-TG2, IgG-DGP,

IgA-DGP and IgA-AGA. POC tests showed a pooled sensitivity of 96.4% for IgA-TG2 (specificity 97.7%).

Conclusions: IgA-EmA and IgA-anti-TG2 tests appear highly accurate to diagnose CD. IgG-anti-DGP tests may help in excluding CD. IgA-AGA and IgA-DGP tests show inferior accuracy. POC tests may achieve high accuracy in the hands of experienced readers, but IgA-anti-TG2/EmA were superior.

Key Words: antibodies, coeliac disease, deamidated gliadin, diagnostic accuracy, point of care tests, systematic review, tissue transglutaminase

(*JPGN* 2012;54: 229–241)

Coeliac disease (CD) is an immune-mediated enteropathy triggered in genetically susceptible individuals by the ingestion of gluten-containing grains (eg, wheat, barley, rye). The disease is associated with human leukocyte antigen (HLA) DQ2 and DQ8 haplotypes and is characterised by intestinal inflammation, villous atrophy, and a relation to autoimmunity. Early diagnosis and dietary treatment can prevent severe, sometimes life-threatening complications (1). Presently existing guidelines from European and North American medical societies for the diagnosis of CD in children include serological testing and the requirement for histological confirmation from small-bowel biopsies (2–4).

With improvement of diagnostic antibody tests, a diagnosis of CD without intestinal biopsies has been suggested in defined situations (5). Omitting biopsies would reduce the burden of endoscopy and of general anaesthesia for the affected children, save costs, and avoid potential adverse effects of these procedures.

The intention of this systematic review and meta-analysis was to update the evidence report on serological tests for CD published in 2004 by the Agency for Healthcare Research and Quality (AHRQ) (6). In the present report, publications from January 1, 2004 to September 1, 2009 were analysed if they used serologic tests in children (anti-gliadin [AGA], endomysial [EmA], anti-transglutaminase-2 [TG2] antibodies as index tests compared with duodenal biopsy results used as the reference standard). In addition to the AHRQ report (6), point-of-care (POC) tests and tests based on deamidated gliadin peptide (DGP) antibodies used as index tests were considered if they were validated against histology as reference. The update was intended to search especially for situations with highly elevated titres where biopsy may be omitted in the diagnosis of CD.

PATIENTS AND METHODS

The protocol of this review was approved by the ESPGHAN Working Group on Coeliac Disease Diagnosis and adhered to current methodological guidelines on the conduct of systematic reviews of diagnostic accuracy (7).

Received January 27, 2011; accepted January 29, 2011.

From the *Centre for Social Policy Research, University of Bremen, Bremen, the †Medical Review Board of the Statutory Health Insurance Funds (MDS), Essen, Germany, the ‡Epidemiology and Biostatistics Unit, Institute for Maternal and Child Health IRCCS Burlo Garofolo, Trieste, Italy, the §Dr von Haunersches Kinderspital, Ludwig Maximilians Universität München, München, Germany, the ||Hans Christian Andersen Children's Hospital, Odense University Hospital, Odense C, Denmark, the ¶Department of Paediatrics, Medical and Health Science Centre, University of Debrecen, Debrecen, Hungary, and the #Heim Pal Children's Hospital, Budapest, Hungary.

Address correspondence and reprint requests to Klaus Giersiepen (e-mail: giersiepen@zes.uni-bremen.de).

The ESPGHAN Working Group on Coeliac Disease Diagnosis coauthors: David Branski, Carlo Catassi, Steffen Husby, Sibylle Koletzko, Ilma R. Korponay-Szabó, Markku Maki, M Luisa Mearin, Alan Phillips, Carmen Ribes-Koninckx, Ranaan Shamir, Riccardo Troncone, Alessandro Ventura, Klaus-Peter Zimmer.

The study was funded by the European Society for Paediatric Gastroenterology, Hepatology and Nutrition. Deutsche Zöliakiegesellschaft, Deutsche Gesellschaft für Kinder- und Jugendmedizin, Coeliac UK, The Danish Coeliac Society, The Spanish Society for Paediatric Gastroenterology.

Copyright © 2012 by European Society for Pediatric Gastroenterology, Hepatology, and Nutrition and North American Society for Pediatric Gastroenterology, Hepatology, and Nutrition

DOI: 10.1097/MPG.0b013e318216f2e5

Eligibility Criteria

We used the search algorithm as AHRQ (6; search details reported in Appendix B of that review) and included cohort studies that enrolled symptomatic children with true diagnostic uncertainty as well as case-control studies of patients with CD (cases) and children with other diseases or relatives of patients with CD (controls, see Table 1). Tests of interest were the measurement of IgA- and IgG-AGA, EmA, anti-TG2, (also named anti-tTG), and—in addition to AHRQ—anti-DGP antibodies as well as POC for IgA-anti-TG2 not requiring laboratory support.

Eligible studies had to report on results of duodenal biopsies for at least 90% of patients classified by Marsh criteria (8) or a corresponding classification allowing assessment of the degree of villous atrophy of duodenal mucosa. Numbers for 2*2 tables with a minimum of $n > 10$ children with CD had to be available to calculate sensitivity, specificity, positive and negative likelihood ratios (“LR+”, “LR-”), and diagnostic odds ratios (DOR), and $>80\%$ of patients with CD had to have at least Marsh II lesions. Median age had to be <18 years. All of the children had to be on a gluten-containing diet. Serological test results were not to be part of the CD case definition to distinguish index tests from reference standard. Only full-text publications in English were included.

More detailed information on 8 exclusion criteria (E1–E8) are provided in Figure 1 and the Appendix. Studies were eliminated in a hierarchical manner, so a study excluded for E1 was not checked further for exclusion criteria E2–E8.

Study Identification and Study Quality Assessment

Two reviewers working independently and in duplicate reviewed all of the abstracts and titles and, upon retrieval of potentially eligible studies, the full-text publications for eligibility. Disagreements were resolved by consensus. Studies were eliminated if they were clearly unrelated to the projects’ topic, judged on review of the title, abstract, and keywords (level 1 screening) or dealt with pathology other than CD, had study objectives other than diagnosis, for example, screening of CD, measuring prevalence of CD, or follow-up of patients with CD (level 2 screening).

Study quality was assessed independently by 2 investigators (K.G., N.S.) according to the Quality Assessment of Studies of Diagnostic Accuracy System (QUADAS (9)). Disagreements were resolved by consensus. Whereas QUADAS reports for 14 items qualitatively, we quantified them by a score for sensitivity analysis to obtain 1 additive summary score per study. Confirmed items were counted as “1,” nonconfirmed items were coded as “-1,” and “unknown” items were handled as “0” (10).

Five QUADAS items were always accepted as present (score +1), as they were part of the inclusion criteria:

- (3) reference standard (in our case: biopsy) likely to correctly classify disease
- (5) whole sample or random sample receive reference standard
- (6) same reference standard regardless of index test result
- (7) index test not part of reference standard
- (12) same clinical data available in study as if in practice

The maximum score could thus be +14 and the minimum score could be -9.

Results of Study Identification

After deletion of duplicates (88), 2510 records were retrieved and entered level 1 screening (Fig. 2), 323 articles were checked at

level 2 and 316 full texts obtained; 87 articles entered level 3 screening, of which $n = 71$ were excluded in a hierarchical manner by exclusion criteria “E1–E8” (Figs. 1, 2).

Data Extraction

All of the eligible studies were abstracted in a standardised form, reporting the study’s goal, design, publication type, study population, median and range of age or mean with standard deviation, proportion female, country, Marsh stage for biopsy (reference test for disease status), checks for presence of IgA deficiency, information on HLA test results in conflicting patients, information to set up 2*2 tables for each index test and subgroup, type, and cutoff of tests. All of the studies reported on more than 1 index test, and if multiple cutoffs were reported, this information was collected in a new dataset, leading to up to 12 datasets per study and a total of 80 datasets for the 16 studies (10–26).

Statistical Analysis

We used MetaDiSc 1.4 software Meta-analysis of Diagnostic and Screening Tests for pooling data (27). Because sensitivities and specificities are interrelated, we also present single and pooled LR and DOR for each index test as summary measures and give information on statistical heterogeneity and the appropriateness of pooling for each test and test parameter.

For sensitivities and specificities for the cutoff given by the manufacturer, 95% confidence intervals (CIs) were calculated with MetaDiSc using the F distribution, and summary estimates were obtained by weighted averages in which the weight of each study is its sample size.

Tests with a high LR+ indicate tests that can help rule in CD, and tests with a low LR- can help rule out CD: an LR+ >10 provides “conclusive evidence” for the disease under study, whereas a low negative likelihood ratio (LR- <0.1) is “conclusive evidence” against the diagnosis (28). Ranges of $5 < LR+ \leq 10$ and $0.1 \leq LR- < 0.2$ give “strong diagnostic evidence.”

DORs describe the ratio of the odds of a positive test result in patients with disease compared with patients without disease and can be calculated as the ratio of the LRs for a positive and a negative test. DORs have the advantage of being a single indicator of test performance providing a global meaning of agreement between a test and a reference standard.

LRs and DORs were pooled with MetaDiSc using a random effects model (27,29); 95% CIs were pooled by

$$LR e^{\pm z_{\alpha/2} SE(\ln LR)}; DOR e^{\pm z_{\alpha/2} SE(\ln DOR)}$$

Cells with 0 content were filled with 0.5. We assessed the inconsistency among studies with MetaDiSc using the inconsistency-squared (I^2) statistic, which represents the proportion of variability across studies that is not the result of chance. I^2 values of $\approx 25\%$, $\approx 50\%$, and $\approx 75\%$ indicate low, moderate, and high heterogeneity, respectively (30).

In case of elevated heterogeneity ($I^2 > 50\%$), pooled results for sensitivity or specificity are not shown. Instead, numbers of studies reaching $\geq 90\%$ and, in some instances, $\geq 95\%$ of sensitivity or specificity are reported, as are ranges.

For sensitivity analysis, subgroups were classified by either biopsy criteria (“Marsh I/II sufficient for CD” vs “Marsh III required for CD”), type of study (cohort vs case-control study), and overall study quality (QUADAS score ≥ 9 vs <9) for the most frequently applied test: IgA-anti-TG2 enzyme-linked immunosorbent assay (ELISA), performed on 15 study populations.

TABLE 1. Study features and populations of included studies

1 Study	2 No. CD patients, median age, age range	3 No. non-CD patients, median age, age range	4 Marsh class no. biopsied (proportion)	5 Tests studied as indicated by authors, company providing ELISA test, numbers refer to text below table	6 Target group, type of study	7 Study quality QUADAS score
Agardh et al (11)	73 5.3 0.7–15.3	50 3.2 0.9–17.3	III 123/123	IgA/IgG RBA/PasRBA anti-tTG Celikey (5); EmA-ME	Suspected CD, cohort	6
Agardh (12)	119 5.7 0.7–19.0	57 3.5 0.9–14.6	III 176/176	IgA-DGP/tTG (Quanta Lite h-tTG/DGP Screen) IgA-DGP (Quanta Lite Gliadin IgA II) IgG-DGP (Quanta Lite Gliadin IgG II) IgA-DGP (Quanta Lite Celiac DGP Screen) IgA-tTG (Quanta Lite h-tTG IgA) IgG-tTG (Quanta Lite h-tTG IgG), all tests provided by (3)	Suspected CD, cohort	4
Barker et al (13)	58 7.7±4.11	45 10.0±4.9	II 103/103	IgA Quanta Lite guinea pig tTG (2) IgA Quanta Lite human tTG (3)	Suspected CD, cohort	10
Basso et al (14)	161 6 1–17	129 9 1–17	I 290/290	IgA-DGP (Quanta Lite Gliadin IgA II) IgG-DGP (Quanta Lite Gliadin IgG II) IgA-tTG IgA (Quanta Lite h-tTG IgA), all tests provided by (3)	Suspected CD, cohort	3
Bazzigaluppi et al (15)	143 Mean 8.8±6.1	64 Mean 6.8±4.8	II 207/207	IgA-EmA-ME, IgA- and IgG-tTG RBA	Cases: suspected CD; controls: relatives of CD patients, or with IDDM, Turner or Down syndrome	6
Blesa-Baviera et al (16)	65–112 3.73 y 11–190 mo	103–113 4.79 y 9–180 mo	III 185/185	Ig-AGA CAP gliadin IgA FEIA IgA-AITGA immunostick (POC, 4) IgA-AITGA/AGA immunostick (POC, 4) IgA-anti-tTG Celikey (5)	Cases: suspected CD; controls: other disease controls	6
Bonamico et al (17)	39 Mean 11.5±11.1	32 Mean 8.1±3.6	IIIb 71/71	Serum tTG (RBA) Saliva tTG (RBA) EmA-ME	Cases: suspected CD; controls: other disease controls	7
Collin et al (18)	16 <2 y 139 No median 0.9–13 y	5 <2 y 121 No median 0.9–13 y	III 21/21 IIIb 260/260	IgA-anti-tTG Celikey (5), EmA IgA-anti-tTG Celikey (5) IgA-AGA (homemade) ELISA or CAP gliadin IgA FEIA Operon Sticks (POC, 4); IgA AITGA/AGA IgA-AITGA	Cases: suspected CD; controls: other disease controls Suspected CD, cohort	12 6
Korponay-Szabó et al (20) “new” patients	38 13* 1.2–72* *Not distinguished by disease status	82 13* 1.2–72* *Not distinguished by disease status	III 120/120	IgA-anti-tTG Celikey (5) IgA-EmA-ME Immunostick IgA (POC, no brand)	Suspected CD, relatives of CD patients or with IDDM; cohort	12
Korponay-Szabó et al (20) “old” patients	99 10.2 1.4–59	65 14.7 3.3–67	III 164/164	IgA-anti-tTG Celikey (5) IgA-EmA-ME Immunostick IgA (POC, no brand)	Suspected CD, cohort	12

(continued)

TABLE 1. (Continued)

1	2	3	4	5	6	7
Study	No. CD patients, median age, age range	No. non-CD patients, median age, age range	Marsh class no. biopsied (proportion)	Tests studied as indicated by authors, company providing ELISA test, numbers refer to text below table	Target group, type of study	Study quality QUADAS score
Lagerqvist et al (21)	428 16 mo	129 37 mo	III 557/557	IgA-AGA Gliadin IgA VARELISA (5) IgA-tTG Celikey (5)	Cases: established CD from Swedish CD registry; controls: other disease controls	10
Llorente et al (22)	7.5 mo–14 y 24; ≤3 y: for AGA and tTG and EmA 14 for EmA	8.5 mo–14.6 y 19; ≤3 y: for AGA and tTG 24 for EmA	III 43/43	IgA-EmA-ME Ig-AGA, IgA-anti-tTG Celikey (5), EmA-ME	Suspected CD and relatives of CD patients, cohort	10
Poddar et al (23)	180	126 Mean 6.3 ± 3.6 range: n.r.	III 306/306	Ig-AGA and IgA-anti-tTG GP; done in 80 CD patients and 40 non-CD-patients (6), EmA-ME in 306	Suspected CD, cohort	10
Prause et al (24)	142	160 Mean 9 range: n.r.	II 302/302	IgA and IgG: AGA, GAF-3X; anti-tTG, Euroimmun (2); DGP; Quanta Lite Gliadin IgA/G II, (5)	Cases: suspected CD; controls: other disease controls	6
Raivio et al (25)	81 12	59 15 1.6–68 for pat. <16	III 140/140	Biocard Celiac Disease (1, POC) EmA-ME IgA-tTG Celikey (5)	Cases: suspected CD or relatives of coeliac patients; controls: other disease controls	9
Yachha et al (26)	23 11 3–18	19 9 5–15	III 42/42	IgA and IgG: AGA (6) IgA-tTG (6) EmA-ME	Cases: suspected CD; controls: other disease controls	13

Studies sorted by first author, number of study participants (columns 2 and 3), reference test criteria (Marsh 1992), tests studied with information on point-of-care tests (POC) and commercial provider of the test, the target group, and the study quality according to the QUADAS score. The study by Korponay-Szabó et al (20) is included with 2 different cohorts: "new", prospective patients and "old", retrospective patients (frozen blood). Anti-tTG corresponds to anti-tTG, IDDM, insulin-dependent diabetes mellitus; ME: monkey oesophagus. Providers of tests: 1. AntiBiotech, Vantaa, Finland. 2. Euroimmun Medizinische Labordiagnostika AG, Lübeck, Germany. 3. INOVA Diagnostics, San Diego, CA. 4. Operon, Zaragoza, Spain. 5. Phadia, Freiburg, Germany. 6. The Binding Site, Schwetzingen, Germany.

1. Biopsy requirement criteria not met	n = 11
2. Age: reported mean or median ≥ 18 y	n = 30
3. Study with ≤ 10 children n=5	
4. "Target group": study did not report on symptomatic children	n = 10
5. No symptomatic reference group with biopsy information	n = 7
6. At least 1 antibody test was part of the case definition	n = 6
7. Test results showed high normal thresholds, assuming laboratory error	n = 1
8. None of the presented index tests commercially available or RIA	n = 1

FIGURE 1. n = 87 studies hierarchically excluded during level 3 screening.

Sensitivities and specificities were pooled for each subgroup and a bivariate χ^2 was calculated, using the Yates correction (EpiInfo statcalc.exe, available at: www.cdc.gov/epiinfo/downloads.htm).

RESULTS

Table 1 summarises the baseline characteristics of the 16 eligible studies: A total of 1876 biopsy-verified patients with CD were compared with 1234 patients without CD from 10 publications of 11 cohort studies (2 cohorts in Korponay-Szabó et al (20) and 6 case-control studies). The numbers of patients with CD per study varied between n = 16 and n = 428 and those of comparison patients between n = 5 and n = 160. If studies presented separate results for children, only data from children were taken, as grouped by the authors, sometimes limited to only a small number of children

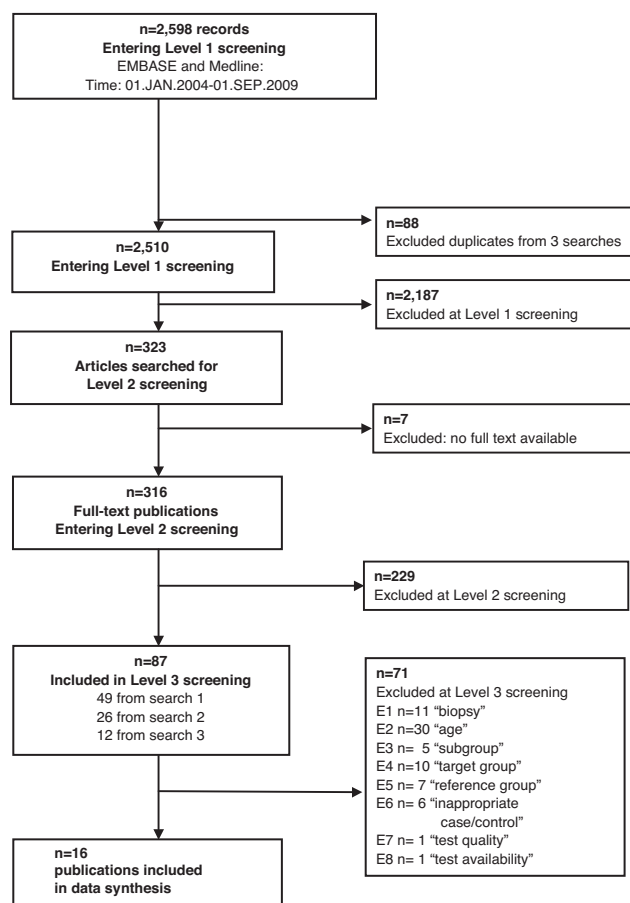


FIGURE 2. Flowchart of study selection by 3 levels of screening and 8 types of exclusion criteria (E1–E8) for level 3 screening.

contributing information: 15 patients with CD younger than 2 years and 5 controls of the respective age were included from Collin et al (18). Llorente et al (22) set the limit at younger than 3 years, yielding 24 patients with CD and 19 comparisons; Raivio et al (25) at 16 years with 81 patients with CD and 59 comparison patients. For the remaining studies, median ages are given or means as well as ranges or standard deviations for age, as reported by the respective studies.

Basso et al (14) had 8/161 patients with duodenal biopsies classified as Marsh I ("infiltrative"), whereas the remaining cases had Marsh II/III ("hyperplastic"/"villous atrophy"). For the populations analysed by Barker et al (13), Bazzigaluppi et al (15), and Prause et al (24), at least Marsh II was required for a CD case, whereas in the remaining studies, at least partial villous atrophy had to be present, corresponding to Marsh IIIa (8). The numbers of patients biopsied per group are reported as well as the reported index tests with information on the test manufacturer and on study quality assessed by the QUADAS summary score.

Data Synthesis

For each set of tests, pooled results are presented in Table 2 for sensitivity, specificity, LR+, LR-, and DOR with ranges and 95% confidence limits, and information on statistical heterogeneity (P value, inconsistency I^2).

IgA-AGA

Six of the included studies reported on IgA-AGA ELISAs, showing a larger range for sensitivity (60.9%–96.0%) than for specificity (79.4%–93.8%). Pooling was not done for sensitivity because of statistically relevant heterogeneity. For 4 studies, both sensitivity and specificity were <90%; Prause presented a low sensitivity (73.9%) and a higher specificity (91.9%), whereas Lagerqvist (2008) reported both a high sensitivity (96.0%) and specificity (93.8%). Pooled specificity was 89.8% (95% CI 86.9%–92.3%).

In the subgroup of children younger than 18 months, Lagerqvist (2008) reported a high sensitivity of 97.2% and a lower specificity of 83.2%.

Pooled LR+ was 7.3 (95% CI 4.5–11.8), pooled LR– 0.186 (95% CI 0.095–0.362), and the DOR 40.6, all 3 pooled estimates showing large heterogeneity, indicated by parentheses in Table 2.

IgG-AGA

Two studies reported on IgG-AGA, reaching sensitivities and specificities of 88.0% and 80.0% (24) and 73.9% and 26.3% (26), respectively (not shown in Table 2).

IgA-anti-TG2: ELISAs and Radiobinding Assays (RBAs)

IgA-anti-TG2 test results were reported from 13 studies plus the 2 cohorts reported by Korponay-Szabó et al (20) on ELISAs (1694 patients with CD and 1138 comparison patients) and from 3 studies on RBAs (255 patients with CD and 146 comparisons).

IgA-anti-TG2: ELISAs

Pooling of studies for sensitivity and specificity was not done because of elevated heterogeneity, as can be seen from Forest plots (Figs. 3 and 4). Ranking the studies by results for sensitivity, 12 out

TABLE 2. Test parameters for 16 studies and pooled results

Test	No. studies	CD patients	Non-CD patients	Test parameter	Range	Pooled estimate	95% Confidence limit	Heterogeneity	
								<i>P</i>	<i>I</i> ²
IgA-AGA (16,19,21,22,24,26)	6	859	511	Sens.	60.9%–96.0%	Heterogenic	Heterogenic	0.000	93.7%
				Spec.	79.4%–93.8%	89.9%	86.9%–92.3%	0.076	49.1%
				LR+	3.6–15.5	(7.3)	4.5–11.8	0.01	66.9%
				LR–	0.04–0.44	(0.186)	0.095–0.362	0.000	92.7%
				DOR	11–366	(40.6)	14.1–117.1	0.000	86.9%
IgA-anti-TG2 ELISAs only: done in all studies except (15) and (17)	15*	1694	1138	Sens.	73.9%–100%	Heterogenic	Heterogenic	0.000	86.0%
				Spec.	77.8%–100%	Heterogenic	Heterogenic	0.000	73.5%
				LR+	4.3–160	(22.4)	12.1–41.4	0.000	74.3%
				LR–	0.01–0.28	(0.06)	0.03–0.10	0.000	85.6%
				DOR	64–19,397	(508)	247–1042	0.007	54.1%
IgA-anti-TG2 RBAs only: (11,15,17)	3	255	146	Sens.	89.0%–100%	Heterogenic	Heterogenic	0.016	75.7%
				Spec.	94.0%–100%	95.9%	91.3%–98.5%	0.21	36.2%
				LR+	14.8–65.2	19.2	9.1–40.6	0.6	0.0%
				LR–	0.013–0.18	(0.06)	0.022–0.165	0.059	64.7%
				DOR	127–5135	347	79–1529	0.15	47.8%
IgG-anti-TG2 ELISAs and RBAs: (11,12,15,24)	4	477	331	Sens.	12.6%–99.3%	Heterogenic	Heterogenic	0.000	99.1%
				Spec.	86.3%–100%	Heterogenic	Heterogenic	0.001	82.2%
				LR+	4.6–21.2	8.3	3.98–17.4	0.12	49.4%
				LR–	0.007–0.88	(0.22)	0.048–0.97	0.000	99.5%
				DOR	5.9–2887	(53.3)	6.9–411	0.000	86.6%
IgA-TG2 POC (16,19,20,25)	5*	470	399	Sens.	94.7%–98.8%	96.4%	94.3%–97.9%	0.56	0.0%
				Spec.	96.6%–98.6%	97.7%	95.8%–99.0%	0.91	0.0%
				LR+	29.1–68.7	40.6	21.3–77.4	0.92	0.0%
				LR–	0.013–0.054	0.04	0.025–0.64	0.65	0.0%
				DOR	1008–2280	1343	547.4–3294	0.95	0.0%
IgA-EmA (11,15,17,18,20–23,25,26)	11*	1034	558	Sens.	82.6%–100%	Heterogenic	Heterogenic	0.000	79.0%
				Spec.	94.7%–100%	98.2%	96.7%–99.1%	0.23	21.9%
				LR+	10.2–160	31.8	19–893	0.86	0.0%
				LR–	0.006–0.18	(0.067)	0.038–0.118	0.000	73.3%
				DOR	64–19397	553.6	219–1402	0.067	42.4%
IgA-DGP (12,14,24)	3	422	346	Sens.	80.7%–95.1%	Heterogenic	Heterogenic	0.000	81.6%
				Spec.	86.3–93.1%	90.7%	87.8%–93.1%	0.14	45.0%
				LR+	6.9–12.7	9.4	6.8–13.1	0.26	25.3%
				LR–	0.06–0.21	(0.121)	0.072–0.203	0.003	78.5%
				DOR	56–93	86.1	55.9–132.7	0.6	0.0%
IgG-DGP (12,14,24)	3	422	346	Sens.	80.1–98.6%	Heterogenic	Heterogenic	0.000	94.4%
				Spec.	86.0–96.9%	Heterogenic	Heterogenic	0.063	59.0%
				LR+	6.8–25.8	(13.6)	8.1–22.8	0.078	55.9%
				LR–	0.02–0.21	(0.061)	0.017–0.22	0.000	92.3%
				DOR	115–948	(234)	100–547	0.092	53.4%

For each set of tests, pooled results are presented for sensitivity, specificity, positive LR+, and negative likelihood ratios LR– as well as for diagnostic odds ratios (DOR) with ranges, 95% confidence limits, and information on statistical heterogeneity (*P*) and inconsistency (*I*²). If inconsistency *I*² was >50%, results were not reported for sensitivity and specificity, but marked as “heterogenic.” For LR+, LR– and DOR, heterogenic estimators are shown in brackets, indicating they may be biased. The study by Korponay-Szabó et al (20) has 2 separate study populations and is counted as 2 studies. *P* values refer to heterogeneity chi-square, obtained by MetaDiSc (27).

of 15 study populations reporting on ELISAs had measured sensitivities ≥90%, whereas 9 achieved sensitivities ≥95%. For specificity, 13 studies found results ≥90% and 11 studies ≥95%. Pooled LR+ was 22.4 (95% CI 12.1–41.4) for all 15 studies and LR– 0.06 (95% CI 0.03–0.1); the DOR was 508 (95% CI 247–1042).

Two study populations from India related to the tests based on guinea pig antigen (GP; (23,26)) were reviewed; 1 started the cohort with GP and replaced it later with human antigen (13), whereas the remainder used human (natural or recombinant) antigens only. Sensitivity was discrepant in the 2 Indian studies, whereas 7 of 10 study populations using human recombinant antigen-based tests (all but 1 using the Celikey test from Phadia) showed sensitivities ≥95% whereas 8 of 10 study populations showed specificities ≥97%. The LR+ for the subset using human recombinant antigen-based tests was 28.9 (95% CI 14.2–58.8), the LR– 0.046 (95% CI 0.024–0.087), and the DOR was 784 (95% CI 362–1702).

IgA-anti-TG2

RBAs showed a similar quality of diagnostic parameters: 2 of 3 studies showed both sensitivities and specificities ≥95% (15,17). Agardh et al (11) reported a sensitivity of 89.0% and a specificity of 94.0%. Pooled LR+ was 19.2 (95% CI 9.07–40.6), LR– 0.06 (95% CI 0.02–0.17), and DOR was 347 (95% CI 78.6–1529).

IgG-anti-TG2: ELISAs and RBAs

There were 4 studies reporting on IgG-anti-TG2. Tests were either ELISAs (12,24) or radioimmunoassay (11,15). Results were heterogeneous; 2 tests showed <30% sensitivity, whereas specificity was ≥94% in 3 studies. The radioimmunoassay used by Bazzigaluppi et al (15) showed the best diagnostic accuracy among IgG-anti-TG2 tests (99.3% sensitivity and 95.3% specificity).

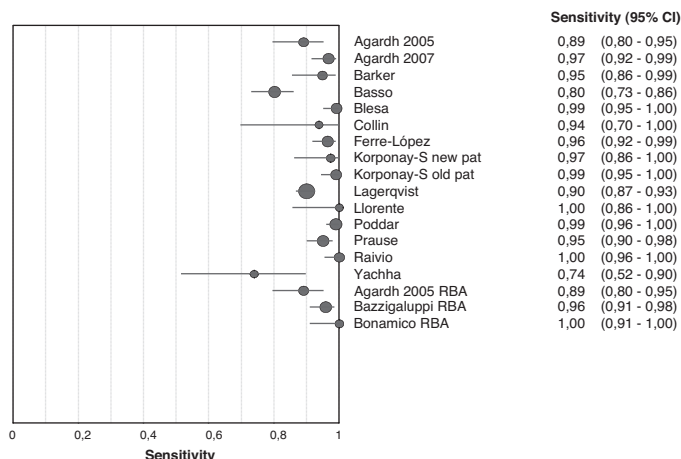


FIGURE 3. Forest plot: sensitivities for IgA-anti-TG2 tests: 15 enzyme-linked immunosorbent assays (top) and 3 radiobinding assays (bottom).

IgA-EmA

IgA-EmA tests were reported from 9 studies plus those from the 2 cohorts by Korponay-Szabó et al (20). For 7 of 11 study populations, the reported sensitivities were $\geq 90\%$ (pooling not done because of too much heterogeneity), and for all 11 studies, specificities were at least 94.7%, yielding a pooled estimate of 98.2%. Pooled LR+ was 31.8, LR- 0.067, and the DOR was 554 (more details shown in Table 2). There were no studies reporting on IgG EmA.

POC Tests: IgA-AGA

POC tests for IgA-AGA were analysed by 2 studies, but results were only shown in combination with anti-TG2. Pooling of POC combination tests for IgA-AGA/anti-TG2 was not done for sensitivity ((16): 95.6% and (19): 99.3%) because of high heterogeneity, whereas pooled specificity was 95.7% (95% CI 91.6%–98.1%) with no substantial heterogeneity (not shown in Table 2).

POC Tests: IgA-anti-TG2

IgA-anti-TG2 POC tests were reported from 4 studies including 2 cohorts by Korponay-Szabó et al (20). For all 5 study populations, sensitivities were $\geq 94.7\%$ (pooled 96.4%; 4 studies $\geq 95\%$); and specificities were all $\geq 96.6\%$ (pooled 97.7%); pooled LR+ was 40.6, LR- was 0.04, and DOR was 1343 (data are presented in more detail in Table 2).

IgA DPG

Results were reported from 3 studies reporting on Quantalite Gliadin IgA II (12,14,24) and a second test used by Prause et al (GAF-3X). Sensitivities were in the range of 80.7% to 95.1%. Two tests reached $\geq 90\%$ sensitivity; 1 passed the threshold $\geq 95\%$.

Specificity ranged between 86.3% and 93.1%, with 3 of 4 results reaching $\geq 90\%$, LR+ was 9.4, LR- 0.12, and DOR was 86.1 (data are presented in more detail in Table 2).

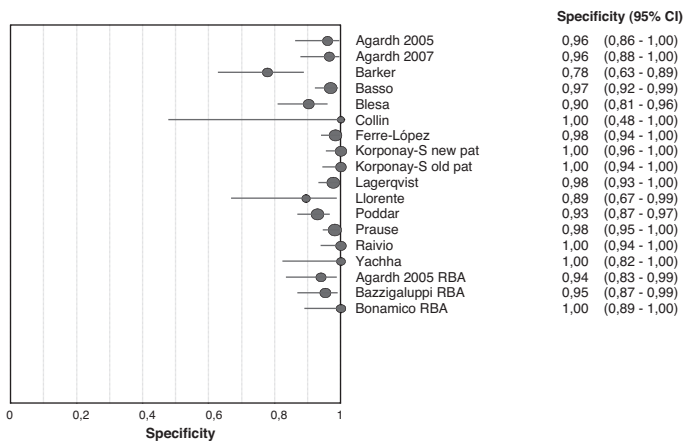


FIGURE 4. Forest plot: specificities for IgA-anti-TG2 tests: 15 enzyme-linked immunosorbent assays (top) and 3 radiobinding assays (bottom).

IgG-DPG

With regard to IgG-DGP, neither sensitivities nor specificities could be pooled from the studies by Basso et al (14), Agardh (12), and Prause et al (24) because of heterogeneity. Sensitivity ranged between 80.1% and 98.6%, the latter 2 reports reaching $\geq 95\%$. As for specificity, 3 out of 4 results were $\geq 90\%$ (all except (12)), whereas the specificity reported by Basso et al (14) surpassed 95%. LR+ was 13.6, LR- was 0.061, and DOR was 234 (data are presented in more detail in Table 2).

Study Quality

The average QUADAS score of the included studies was 8.1 (median 8, range 3–13; a complete item description and individual study results can be found in (10)).

Sensitivity analyses (for IgA-anti-TG2 only)

Studies requiring histological Marsh III lesions showed higher test sensitivities for IgA-anti-TG2 (94.6%) than studies that also included patients with Marsh I/II+ lesions (88.4%; $P < 0.001$), whereas for specificity there was no substantial discrepancy (94.9% for Marsh I/II+ for CD vs 96.6% for Marsh III as minimum requirement; $P = 0.25$).

The QUADAS score did not have any impact on the measured sensitivities ($P = 0.9$) nor specificities ($P = 0.5$), nor did the type of study (cohort vs case-control studies; difference for sensitivity: $P = 0.44$, for specificity: $P = 0.25$).

Complete Concordance (100%) of Laboratory Tests With Biopsy Classification

Of 80 test constellations reported by the 16 studies, 7 yielded sensitivities of 100% (“no patient with CD missed by the index test”) and 17 reported specificities of 100% (“no one without CD wrongly classified as patient with CD”). Four test constellations were without any classification error, showing both 100% sensitivity and 100% specificity (2 for IgA EmA, 1 for IgA-anti-TG2 ELISA, and 1 for IgA-anti-TG2 RBA ((17,22,25), respectively).

Studies Reporting on Multiple Cutoffs

Agardh (12), Barker et al (13), and Poddar et al (23) reported results for multiple cutoffs: Agardh and Barker et al both used Quanta Lite by Inova Diagnostics (San Diego, CA). Agardh (12) used the recommended cutoff of 20 U/mL as well as an additional study cutoff at 30 U/mL, whereas Barker et al used 20 and 100 U/mL with the intention to search for thresholds of sufficient diagnostic accuracy to avoid biopsy. The Quanta Lite kit in the Barker et al study had been changed from GP based to human TG2 antigen during the study. Poddar et al (23) reported on the test by The Binding Site at a cutoff of 4 U/mL (the recommended cutoff by the manufacturer) and at 10 and 20 U/mL.

Perfect specificity (no false-positives) was reported by Agardh (12) for IgA-TG2 at a cutoff at 30 U/mL and for IgG at 20 U/mL. Sensitivity was 95% for IgA-TG2 at a cutoff at 30 U/mL, but low for IgG: 12.6% at 20 U/mL and 4.6% at 30 U/mL.

Barker et al (13) improved specificity from 80% to 97.7% for IgA-anti-TG2 at 100 U/mL, wrongly classifying 1 of 45 patients without CD as CD positive, but lost 7 true patients with CD (sensitivity decreased from 94.8% to 82.7%) compared with the cutoff suggested by the manufacturer for the Quanta Lite test.

INTERPRETATION AND DISCUSSION

We conducted a systematic review and meta-analysis of studies that enrolled mostly paediatric patients with suspected CD and performed antibody detection index tests for CD with biopsy as the reference standard test. Our review included studies published after 2003; older studies were previously summarised by AHRQ (6).

Sensitivity is crucial for finding new patients, but for evaluating whether seropositivity for coeliac antibodies would correctly detect CD without performing an intestinal biopsy, specificity is the critical parameter. Sensitivity and specificity are not independent parameters, and in general, by lowering a threshold cutoff of a particular test kit, sensitivity will be increased by concomitant loss of specificity.

Tests With Best Performance: EmA and Anti-TG2

The synthesis of the evaluated 80 datasets in 16 studies indicated that IgA-anti-TG2 and EmA were the best laboratory tests predicting CD. For both tests, LRs were in the range of “conclusive evidence” (28) for either confirming or rejecting a diagnosis of CD in case of a positive (LR+) or negative (LR-) test result (Table 2; test statistics).

Sensitivity of EmA tests ranged lower (7/11 studies presented $\geq 90\%$ sensitivity; 6/11 studies presented sensitivities $\geq 95\%$) compared with IgA-anti-TG2 (12/15 studies on ELISAs presented $\geq 90\%$ sensitivity; 9/15 studies $\geq 95\%$). In contrast, the specificity was more stably higher for EmA (pooled 98.2%, all except 1 study reached $\geq 95\%$ specificity, whereas 11/15 studies on IgA-anti-TG2 reached 95% specificity).

When only IgA-anti-TG2 ELISAs based on human recombinant antigens were analysed, the performance improved particularly regarding sensitivity. All but 1 study used the same commercial test, however, and it is thus unclear whether all human recombinant TG2-based tests would perform similarly well.

IgA-anti-TG2 could work better as an initial CD detection test, whereas the highly specific EmA test could be used as a confirmatory test to identify test positives as true patients with CD.

Comparing Our Results With AHRQ: EmA

AHRQ (6) evaluated 18 studies on EmA tests based on monkey oesophagus substrate and reported a pooled sensitivity of 96.1% (95% CI 94.5%–97.3%); all but 1 study had shown sensitivities $\geq 90\%$; 14 studies yielded sensitivities $\geq 95\%$. In our 11 study populations, there were 1529 patients tested with EmA, the majority (1508) tested with monkey oesophagus as substrate. We could not pool sensitivities because of statistical heterogeneity, but the proportion of studies reaching either 90% or 95% sensitivity was lower than in the AHRQ compilation.

This result is not because of a less stringent reference standard in our sample. Only 1 study reporting on IgA EmA also had included patients with CD with $<$ Marsh III lesions at biopsy (15), but the sensitivity for EmA was not different from the remaining 4 studies, including patients with at least Marsh III findings. Instead, the lower sensitivities observed in our study compilation may be related to the difference in the study populations. In the last decade, more patients are coming to clinician attention with mild clinical symptoms, and the overall proportion of milder histology lesions even within the Marsh III category is relatively higher. It is known that patients with Marsh IIIa grade villous lesion are less frequently EmA positive than those with Marsh IIIb or IIIc (31).

Our finding of a pooled specificity of 98.2% for EmA (95% CI 96.7%–99.1%) was similarly high as in the compilation by AHRQ: 97.4% (95% CI 96.3%–98.2%). The pooled LR+ of 38.1 in our study compilation makes CD likely in a patient with a positive EmA test result.

Comparing Our Results With AHRQ: Anti-TG2

For IgA-anti-TG2 (ELISAs and RBAs), we could not confirm AHRQ's excellent results for anti-TG2 measurements performed with human recombinant antigen. There were only 3 such articles available in 2004, and all of them had shown sensitivities and specificities $\geq 95\%$ for IgA-anti-TG2 (32–34) while in our subset; this target was reached only by 7/10 (for sensitivity) and 8/10 (for specificity) study populations. Our total pool presented 18 test results: 11/18 study populations (all ELISAs and RBAs) reached $\geq 95\%$ sensitivity and 13/18 reached a specificity level $\geq 95\%$; however, the accuracy improved when only ELISAs based on human antigen were considered.

Quality criteria for the reference standard were less stringent, because histological Marsh I/II changes were sufficient for CD diagnosis in 3 of our studies on IgA-anti-TG2. The number of patients was larger in our sample (2,115 patients tested with ELISAs based on human recombinant substrate), while AHRQ reported on 3 studies with a total of 216 patients based on human substrate. This discrepancy is probably related to the fact that at the time of the AHRQ study most anti-TG2 tests were still GP antigen based and human TG2 antigens were just introduced and reported in typical case sets. Since then, human-TG2 antigens became universally available and the use of the human antigens predominant. Other sample sources such as whole blood, saliva and stool were explored as well, but yielded inferior results for saliva and stool (10).

IgA-AGA and IgG-AGA

Six studies reported on IgA-AGA and 2 on IgG-AGA tests, all but 1 showing inferior diagnostic accuracy than EmA or IgA-anti-TG2, which is in concordance with statements by North American Society for Pediatric Gastroenterology, Hepatology, and Nutrition (NASPGHAN) and National Institute for Health and Clinical Excellence (3,4). Lagerqvist et al (21) emphasised their high sensitivity found in Swedish children younger than 18 months, which seems to contradict the recommendation of NASPGHAN (3) to abandon AGA analysis in routine childhood CD testing. Lagerqvist et al concluded that 17% of the children with CD in the youngest age group would have remained undiagnosed if IgA-anti-TG2 had been used alone. Our suggestion is to be cautious with the interpretation of IgA-AGA tests because follow-up with challenge procedures have not been reported in most of these children.

Anti-DGP

Tests based on IgG-DGP yielded a pooled LR+ of 13.6, equivalent to “conclusive evidence” for the presence of disease as classified by Deeks (28), but ranking lower than tests for EmA and IgA-anti-TG2. Detection of IgA-DGP antibodies was found to have lower diagnostic accuracy, as defined by the proportion of correctly identified individuals. Because detection of DGP antibodies was not available at the time of the AHRQ report, there are no data on DGP antibodies to compare our findings.

Future studies will have to show whether the relatively high specificity of IgG-DGP antibodies will be useful, especially in detecting CD in IgA-deficient individuals.

POC

There were results from 5 study populations reporting on POC tests in a total of 869 patients and comparing them with serologic laboratory-based tests: IgA-anti-TG2 (in all 5 study populations) and EmA in 3 study populations (20,25).

For IgA-anti-TG2 ELISA (all 4 studies used Celikey, Phadia), the pooled sensitivity was 98.3% (95% CI 96.7%–99.3%) in the 5 studies reporting on POC tests, whereas the sensitivity of POC tests reached 96.4% (95% CI 94.3%–97.9%; Table 2). It was not possible to pool specificity for IgA-anti-TG2 ELISAs because of heterogeneity in the 5 studies (Blesa-Baviera et al reached 90.3%, whereas all of the others achieved at least 98.3%). The pooled specificity for POC tests was 97.7% (95% CI 95.8%–99.0%; Table 2).

For the 424 patients in this subgroup also tested by EmA, the pooled sensitivity of EmA was 99.1% (95% CI 96.7%–99.9%), and the pooled specificity was 100% (95% CI 98.2%–100%).

For POC tests it can thus be concluded that they had high diagnostic accuracy, but the corresponding IgA-EmA and anti-TG2 done on the same individuals showed even better diagnostic parameters.

In spite of apparently excellent diagnostic accuracy, there is no indication that POC tests are superior to IgA-anti-TG2 or EmA tests. Presumably the studies reporting on POC tests had more homogenous study populations: all of the patients with CD analysed had to have Marsh III on biopsy and 1 of the 2 cohorts in the study by Korponay-Szabó et al (20) comprised patients with CD previously diagnosed as having CD (Fig. 5), so there is probably some preselection of patients with extended signs of CD.

All test results reported in the meta-analysis were based on study populations with a high prevalence of CD: 60.3% of all patients studied in our pool had biopsy-confirmed CD. The positive predictive value in that situation is 98.5% for POC tests, the negative predictive value is 94.7%. Assuming instead a prevalence of CD in 5% of all symptomatic children—more likely to correspond to the situation in a general practitioner's office—the positive predictive value would drop to 68.6% and the negative predictive value would rise to 99.8%.

Although POC tests had been used and interpreted by paediatricians in the 4 reported studies, they may not work well in the hands of laypeople or medical staff with little experience in the interpretation of such tests. The possibility that POC can be performed by laypeople not considering the circumstances at the time of testing such as gluten intake, absence of interfering drugs, or IgA deficiency may further interfere with the reliability of the test results. In addition, POC tests may have the inherent problem that they are currently nonquantitative. Future studies will have to show whether the highly accurate results obtained by specialists will be reproduced in less selected populations and by individuals less experienced in interpreting POC test results.

1. There is no clear statement to be derived from the 16 studies with regard to the omission of biopsy at a particular cutoff for a given laboratory test: only 1 study examined this issue (Barker, 2005).
2. IgA anti-TG2 and EmA are the cornerstones of laboratory-based CD diagnosis.
3. The new tests based on anti-DGP for IgA and for IgG did not have a better diagnostic performance than IgA-anti-TG2 and EmA.
4. The studies reporting on the new POC tests showed much more homogeneous results for both POC and corresponding serologic tests, probably due to more selected study populations within these studies. POC tests are no substitute yet neither for biopsy nor for laboratory-based tests, especially not in the hands of inadequately trained readers.

FIGURE 5. Key statements.

Higher Cutoffs for Increasing Specificity

The present evaluation also investigated whether specificity could be increased to 100%, safely predicting small intestinal villous atrophy in children by raising the cutoffs from laboratory tests. A study in adults by Hill et al reported that human-TG2 antigen-based serum anti-TG2 ELISA values always were associated with villous atrophy when they exceeded 10 times the upper limit of normal for a test calculating antibody concentration from a standard curve (5). In such cases, small intestinal biopsy would be unnecessary. There is a further study published by Dahlbom et al after the end of our literature search showing that similar serum IgA-anti-TG2 levels and test result calculations as reported by Hill et al (5), combined with high IgG-anti-TG2 values, also could predict villous atrophy in children (35).

In our selection of 16 studies, 3 investigated the performance of serum antibody tests at various cutoff levels (Agardh (12): 1.5 times the upper limit of normal cutoff for 6 tests; Barker et al (13): 5 times; and Poddar et al (23): 2.5 times and 5 times the upper limit of normal cutoff, all for IgA-anti-TG2), but none of these tests were standard curve based and may thus give different results in different runs. Agardh (12) reported no false-positives for IgG-anti-TG2 already at the cutoff suggested by the manufacturer, therefore, raising the cutoff could not improve specificity.

Among the 80 datasets we analysed from the 16 studies, there were 17 constellations with a specificity of 100%: 9 sets from IgA-anti-TG2, 6 sets from EmA, and 2 sets from combined test kits using IgA-anti-TG2 and IgG-DGP. Only 1 test situation had changed to 100% by raising the cutoff from the limit suggested by the manufacturer: Agardh (12) increased the specificity from 96.6% to 100% for IgA-anti-TG2 by raising the cutoff 1.5 times (30 U/mL; 2 more patients without CD became correctly identified as nondiseased).

Barker et al (13) correctly classified 44 of 45 biopsy-negative as patients without CD by IgA-anti-TG2 at a cutoff 5 times higher than the upper limit of normal value with the same test as Agardh (12), whereas 9 additional patients were seropositive at the normal cutoff but negative on biopsy. Barker et al suggested that patients with an IgA-anti-TG2 > 5 times the upper limit of normal value should be biopsied only if they do not experience symptomatic improvement with a gluten-free diet.

Neither the anti-TG2 kit used by Agardh (12) (human antigen) nor that by Barker et al (13) (GP or human antigen) is standard curve based, so it is unclear how these cutoff proposals correspond to the cutoffs described by Hill et al and Dahlbom et al (5,35), or to values in other commercial tests. Furthermore, unlike other tests, 100 U/mL in the Barker et al study is not an upper limit of the measurement range and may in reality not represent high values (36).

Dahlbom et al correctly identified all 109 children with histologically confirmed CD by Celikey test, equivalent to 100% sensitivity (35). Among the 130 controls (both children and adults) there was 1 person wrongly positive at the cutoff of 3 U/mL with a finding of <10 U/mL. As communicated to us by 1 of the authors (Korponay-Szabó), this person happened to be a child, so child-specific specificity could be calculated as 99.3%. To avoid even this small proportion of false-positive patients, Dahlbom et al suggested raising the limit to 30 U/mL, 10 times the suggested threshold of 3 U/mL, which is considered superior to the manufacturer's suggested cutoff.

Impact of Study Quality and Reference Standards

There was no appreciable difference in the diagnostic yield between studies with a high or low QUADAS score, so the quality

of the studies was related neither to sensitivity nor to specificity. We did not find a difference in test outcomes for case-control versus cohort studies; this may be interpreted as an argument against a selection of more clearly symptomatic patients in the case of a case-control study.

It was a surprising finding that raising the level of the reference standard from Marsh I/II to Marsh III improved the sensitivity of index tests, as less—only the more severely—diseased patients should then be picked up by the same laboratory test. The 4 studies including individuals without villous atrophy as patients with CD did not report their results as stratified by the Marsh classification, so this issue cannot be evaluated further. QUADAS ranking was on average lower (score 4.5) in those 4 studies than the entire study pool's average (score 8.1), so lower overall study quality may be an explanation for this unexpected finding.

Because our search included only studies where >80% of patients with CD showed a Marsh II or higher grade small intestinal lesion, it cannot be excluded that some of the seropositive subjects regarded as false-positives could in fact be truly gluten-sensitive subjects. In addition to the sometimes slowly evolving nature of the small intestinal lesion, it must be noted that histological lesions similar to those seen in CD also may rarely occur in other diseases.

Limitations and Strengths

Our study was designed to verify intestinal lesions of CD. It is therefore obvious that patients with only extraintestinal forms of CD may have shown positive serology but negative intestinal biopsy, thus pretending a false-positive test result. Such a condition may be, for example, dermatitis herpetiformis, which was not sufficiently represented in the evaluated material.

Some of the case-control studies were not designed to enroll all subsequent patients, but made a selection from a much larger pool. Possibly these patients were neither representative CD cases nor representative controls. Furthermore, some of the patients probably had undergone serological testing before recruitment, leading to an artificially raised prevalence estimate. The pooled prevalence of CD in the cohort study populations was indeed high: 56.2%, making such a bias most likely, although such selections were not reported in the 16 studies.

Histology has been questioned as the predominant diagnostic tool in CD (37–40), because it can be misleading by producing mainly false-negative results. The lesions may be patchy and thus missed if only 1 biopsy is taken from the descending part of the duodenum or by capsule biopsy. The ongoing debate on the validity of histology for CD verification, its exclusive role as reference standard, and the specificity of villous atrophy including intra-epithelial lymphocytosis for CD is acknowledged. In spite of the doubts expressed in those studies and potential technical pitfalls in the pathology interpretation, it was decided to retain biopsy histology as the single reference standard, because the inclusion of information from serologic tests as part of the reference standard definition would have led to circular reasoning making validation impossible.

Both EmA- and TG2-based ELISA tests in principle detect the same antibodies (40), so variability across different studies may be more dependent on the differences in the exposure of the TG2 antigenic sites in a particular test than on patient characteristics.

CONCLUSIONS AND IMPLICATIONS FOR PRACTICE AND RESEARCH

The AHRQ report from 2004 did not make an explicit statement on the necessity of diagnostic biopsy, but the North American guideline concluded, "It is recommended that confirmation of the diagnosis of CD require an intestinal biopsy in all

cases” (3), as did other currently valid guidelines for children (2), and for children and adults (4). Our findings are the basis for the new ESPGHAN guideline published in the January 2011 issue.

The test results for EmA are comparable to the situation as described by AHRQ despite a more complex and more difficult clinical situation than in 2004. For human recombinant IgA-anti-TG2, the new data show variable results and raise some doubt about the excellent diagnostic accuracy presented for these tests in the AHRQ report.

The new POC tests and tests based on DGP antibodies should be followed up in future studies. Although POC tests are performed with high accuracy in expert hands, it is too early to classify them as substitutes for laboratory tests. Tests for IgG-DGP may play a role in the diagnosis of CD in IgA-deficient children in addition to the already established value of IgG-anti-TG2 (41,42).

Neither NASPGHAN (3) nor National Institute for Clinical Excellence (4) had proposed situations when biopsy can be omitted in the diagnosis of CD. We only found 1 small study in children approaching this question by introducing variable cutoffs for laboratory tests, yet ending with inconclusive results (13).

The broader evidence background for the present literature analysis did not disclose higher accuracies in the reported studies despite improved technical approaches. This lack of improvement probably reflects more unselected patient populations. It may be

argued that the tests have stood the test of time and the consequence of such high accuracies could be the omission of histological analysis in selected cases.

Future studies reporting on new diagnostic tests for CD should be designed as cohort studies enrolling consecutive patients, including IgA, EmA, and probably also human leucocyte antigen testing in the reference standard besides biopsy. These studies should include patient follow-up for the detection of potential misclassification.

Conflict of Interest Statements: The following authors declared potential conflicts of interest: S. Koletzko (research support from Euroimmun, Phadia, instrumentation laboratory for 1 research project), I. Korponay-Szabó (patent application on point of care coeliac test licensed by the University of Tampere, Finland, to AniBiotech, Vantaa, Finland), Carlo Catassi (consultant for Menarini Diagnostics s.r.l., Florence, Italy), C. Ribes-Koninckx (research support from Phadia), and M. Maki (consultancies for Finnish Food Safety Authority Evira, The Finnish Funding Agency for Technology and Innovation Tekes; The Finnish Innovation Funds Sitra; International Life Science Institute; Coeliac Research Fund, Australia; Domm International; Finn Medi; Sinebrychoff; Moilas; Raisio; Phadia; Anibiotech; Kustannus Duodecim, Finland; Vac-tech; Eurospital; Inova; Nexpep; Alvine Pharmaceuticals; Shire; GlaxoSmithKline; Alba Therapeutics; ChemoCentryx; Zedira).

APPENDIX A STUDY SELECTION: INCLUSION/EXCLUSION CRITERIA

Short name	Inclusion	Exclusion
E1 Biopsy	Biopsy part of primary diagnosis: Classification system has to be described or Reference to Marsh, Oberhuber, etc must be stated or Wording must contain corresponding information on hyperplastic lesions with enlarged crypts, with increased crypt cell division (corresponding to Marsh II lesion) or “partial/total villous atrophy” (corresponding to Marsh III lesion) >90% of CD patients and >90% comparison patients must have biopsy results presented by classification system >80% of CD patients must have ≥Marsh II lesions	No information on biopsy at time of diagnosis, information on follow-up biopsy only, biopsy only mentioned in the context with “diagnosis according to ESPGHAN criteria 1990 (Walker-Smith)” without further statement on classification system or wording on type of atrophy ≤90% of CD patients or ≤90% of comparison patients have biopsy results presented by classification system ≤80% of CD patients have ≥Marsh II lesions
E2 Age	Age: children only Median or mean age (whatever reported) at time of primary diagnosis must be <18 years for both boys and girls and both groups (CD; control)	Adults only or Age not stated or Age median or mean ≥18 years
E3 Subgroup	Number of children with CD must be n > 10 in studied population Results must be presented so 2*2 tables can be filled to calculate sensitivity, specificity, PPV, and NPV for children Required setting for counting numbers of children at the time of primary diagnosis of CD only, Follow-up exams do not count	Number of children with CD at primary diagnosis n ≤10 or 2*2 tables cannot be constructed for children
E4 Target group	AHRQ target groups 2, 3, or 4 (as defined by AHRQ, page 12/906) have to be discernable: “Group 2”: “symptoms” “Group 3”: “asymptomatic at risk” “Group 4”: “asymptomatic + feature” (eg, type 2 diabetes mellitus)	AHRQ group 1: screening for CD in asymptomatic general population in addition to AHRQ: children with CD, but on gluten-free diet are excluded: follow-up performance on gluten-free diet is not relevant
E5 Reference group	Test results on CD-free reference group must be reported	No potentially CD-free reference group available, no results on CD-free reference group reported
E6 Inappropriate case and/or control group	Test under study not part of CD/no CD classification	If cases had to be positive and/or controls had to be negative for case definition, adult blood donors may not serve as controls
E7 Test quality	Lab test titres should not be influenced by contamination	Contamination likely to explain unusual titres
E8 Test availability	At least 1 reported test must be commercially available or radioimmunoassay	Reported tests are an in-house/research kits only

APPENDIX B CALCULATION OF DIAGNOSTIC PARAMETERS

Sensitivity (sens), specificity (spec), positive and negative likelihood ratios LR+ or LR–, diagnostic odds ratios (DOR), and diagnostic accuracy

Results in an individual study

Total	Reference test		
	CD	Without CD	
Positive	a (TP)	b (FP)	P
Negative	c (FN)	d (TN)	N
Total	D	ND	T

a: Number of people with positive test result and CD: true-positives (TP)
 b: Number of people with positive test result and without CD: false-positives (FP)
 c: Number of people with negative test result and CD: false-negatives (FN)
 d: Number of people with negative result and without CD: true-negatives (TN)

P: Total number of people with positive test result
 N: Total number of people with negative test result
 D: Total number of people with CD (diseased)
 ND: Total number of people without CD (non diseased)
 T: Number of people in the study

Sens true-positives (TP)/D
 Spec true-negatives (TN)/ND
 LR+ true-positives (TP)/false-positives (FP) or: sensitivity/(1-specificity)
 LR– false-negatives (FN)/true-negatives (TN) or: (1-sensitivity)/specificity
 DOR (true-positives [TP]/false-positives [FP])/(false-negatives [FN]/true-negatives [TN]) or: LR+/LR–
 Diagnostic accuracy (true-positives [TP] + true-negatives [TN])/T

REFERENCES

- Fasano A, Bertl I, Gerarduzzi T, et al. Prevalence of celiac disease in at-risk and not-at-risk groups in the United States: a large multicenter study. *Arch Intern Med* 2003;163:286–92.
- Walker-Smith JA, Guandalini S, Schmitz J, et al. Revised criteria for diagnosis of coeliac disease. *Arch Dis Child* 1990;65:909–11.
- Hill ID, Dirks MH, Liptak GS, et al. Guideline for the diagnosis and treatment of celiac disease in children: recommendations of the North American Society for Pediatric Gastroenterology, Hepatology and Nutrition. *J Pediatr Gastroenterol Nutr* 2005;40:1–19.
- National Institute for Clinical Excellence. CG86 coeliac disease: full guideline. <http://guidance.nice.org.uk/CG86/Guidance/pdf/English>. Accessed October 7, 2010.
- Hill PG, Holmes GK. Coeliac disease: a biopsy is not always necessary for diagnosis. *Aliment Pharmacol Ther* 2008;27:572–7.
- Rostom A, Dubé C, Cranney A, et al. Summary, Evidence Report/Technology Assessment no. 104. <http://www.ahrq.gov/clinic/epcsums/ceciacsum.htm>. Accessed October 7, 2010.
- Deville WL, Buntinx F, Bouter LM, et al. Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Med Res Methodol* 2002;2:9.
- Marsh MN. Gluten, major histocompatibility complex, and the small intestine: a molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). *Gastroenterology* 1992;102:330–54.
- Whiting P, Rutjes AW, Reitsma JB. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;3:25.
- Giersiepen K, Leigemann M, Stuhldreher N, et al. Evidence report on serological and point of care tests for the diagnosis of coeliac disease. www.espgan.med.up.pt. Accessed October 7, 2010.
- Agardh D, Dahlbom I, Daniels T, et al. Autoantibodies against soluble and immobilized human recombinant tissue transglutaminase in children with celiac disease. *J Pediatr Gastroenterol Nutr* 2005;41:322–7.
- Agardh D. Antibodies against synthetic deamidated gliadin peptides and tissue transglutaminase for the identification of childhood celiac disease. *Clin Gastroenterol Hepatol* 2007;5:1276–81.
- Barker CC, Mitton C, Jevon G, et al. Can tissue transglutaminase antibody titers replace small-bowel biopsy to diagnose celiac disease in select pediatric populations? *Pediatrics* 2005;115:1341–6.
- Basso D, Guariso G, Fogar P, et al. Antibodies against synthetic deamidated gliadin peptides for celiac disease diagnosis and follow-up in children. *Clin Chem* 2009;55:150–7.
- Bazzigaluppi E, Roggero P, Parma B, et al. Antibodies to recombinant human tissue-transglutaminase in coeliac disease: diagnostic effectiveness and decline pattern after gluten-free diet. *Dig Liver Dis* 2006;38:98–102.
- Blesa-Baviera LC, Donat-Aliaga E, Ortigosa L, et al. Celiac disease screening by immunochromatographic visual assays: results of a multicenter study. *J Pediatr Gastroenterol Nutr* 2007;45:546–50.
- Bonamico M, Ferri M, Nenna R, et al. Tissue transglutaminase autoantibody detection in human saliva: a powerful method for celiac disease screening. *J Pediatr* 2004;144:632–6.
- Collin P, Kaukinen K, Vogelsang H, et al. Antiendomysial and antihuman recombinant tissue transglutaminase antibodies in the diagnosis of coeliac disease: a biopsy-proven European multicentre study. *Eur J Gastroenterol Hepatol* 2005;17:85–91.
- Ferre-Lopez S, Ribes-Koninckx C, Genzor C, et al. Immunochromatographic sticks for tissue transglutaminase and anti-gliadin antibody screening in celiac disease. *Clin Gastroenterol Hepatol* 2004;2:480–4.
- Korponay-Szabó IR, Raivio T, Laurila K, et al. Coeliac disease case finding and diet monitoring by point-of-care testing. *Aliment Pharmacol Ther* 2005;22:729–37.
- Lagerqvist C, Dahlbom I, Hansson T, et al. Antigliadin immunoglobulin A best in finding celiac disease in children younger than 18 months of age. *J Pediatr Gastroenterol Nutr* 2008;47:428–35.
- Llorente MJ, Sebastian M, Fernandez-Acenero MJ, et al. IgA-antibodies against tissue transglutaminase in the diagnosis of celiac disease: concordance with intestinal biopsy in children and adults. *Clin Chem* 2004;50:451–3.
- Poddar U, Thapa BR, Nain CK, et al. Is tissue transglutaminase autoantibody the best for diagnosing celiac disease in children of developing countries? *J Clin Gastroenterol* 2008;42:147–51.

24. Prause C, Ritter M, Probst C, et al. Antibodies against deamidated gliadin as new and accurate biomarkers of childhood coeliac disease. *J Pediatr Gastroenterol Nutr* 2009;49:52–8.
25. Raivio T, Kaukinen K, Nemes E, et al. Self transglutaminase-based rapid coeliac disease antibody detection by a lateral flow method. *Aliment Pharmacol Ther* 2006;24:147–54.
26. Yachha SK, Aggarwal R, Srinivas S, et al. Antibody testing in Indian children with coeliac disease. *Indian J Gastroenterol* 2006;25:132–5.
27. Zamora J, Abaira V, Muriel A, et al. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol* 2006;6:31.
28. Deeks JJ. Systematic reviews of evaluations of diagnostic and screening tests. In: Egger M, Smith GD, Altman DG, eds. *Systematic Reviews in Health Care. Meta-analysis in Context*. London: BMJ Books; 2001: 248–82.
29. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177–88.
30. Higgins JPT, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. *Br Med J* 2003;327:557–60.
31. Rostami K, Kerckhaert J, Tiemessen R, et al. Sensitivity of antiendomysium and antigliadin antibodies in untreated coeliac disease: disappointing in clinical practice. *Am J Gastroenterol* 1999;94:888–94.
32. Hansson T, Dahlbom I, Hall J, et al. Antibody reactivity against human and guinea pig tissue transglutaminase in children with coeliac disease. *J Pediatr Gastroenterol Nutr* 2000;30:379–84.
33. Wolters V, Vooijs-Moulaert A, Burger H, et al. Human tissue transglutaminase enzyme linked immunosorbent assay outperforms both the guinea pig based tissue transglutaminase assay and anti-endomysium antibodies when screening for coeliac disease. *Europ J Pediatrics* 2002;161:284–7.
34. Vitoria JC, Arrieta A, Ortiz L, et al. Antibodies to human tissue transglutaminase for the diagnosis of coeliac disease. *J Pediatr Gastroenterol Nutr* 2001;33:349–50.
35. Dahlbom I, Korponay-Szabó IR, Kovács JB, et al. Prediction of clinical and mucosal severity of coeliac disease and dermatitis herpetiformis by quantification of IgA/IgG serum antibodies to tissue transglutaminase. *J Pediatr Gastroenterol Nutr* 2010;50:140–6.
36. Van Meensel B, Hiele M, Hoffman I, et al. Diagnostic accuracy of ten second-generation (human) tissue transglutaminase antibody assays in coeliac disease. *Clin Chem* 2004;50:2125–35.
37. Brown I, Mino-Kenudson M, Deshpande V, et al. Intraepithelial lymphocytosis in architecturally preserved proximal small intestinal mucosa: an increasing diagnostic problem with a wide differential diagnosis. *Arch Pathol Lab Med* 2006;130:1020–5.
38. Corazza GR, Villanacci V. Coeliac disease. *J Clin Pathol* 2005;58:573–4.
39. Corazza GR, Villanacci V, et al. Comparison of the interobserver reproducibility with different histologic criteria used in coeliac disease. *Clin Gastroenterol Hepatol* 2007;5:838–43.
40. Korponay-Szabó IR, Laurila K, Szondy Z, et al. Missing endomysial and reticulin binding of coeliac antibodies in transglutaminase 2 knock-out tissues. *Gut* 2003;52:199–204.
41. Cataldo F, Lio D, Marino V, et al. IgG(1) antiendomysium and IgG antitissue transglutaminase (anti-tTG) antibodies in coeliac patients with selective IgA deficiency. Working Groups on Coeliac Disease of SIGEP and Club del Tenue. *Gut* 2000;47:366–9.
42. Korponay-Szabó IR, Dahlbom I, Laurila K, et al. Elevation of IgG antibodies against tissue transglutaminase as a diagnostic tool for coeliac disease in selective IgA deficiency. *Gut* 2003;52:1567–71.