



LUND UNIVERSITY

Validity of histology for the diagnosis of paediatric coeliac disease: a Swedish multicentre study.

Montén, Caroline; Bjelkenkrantz, Kaj; Gudjonsdottir, Audur H; Browaldh, Lars; Arnell, Henrik; Naluai, Åsa Torinsson; Agardh, Daniel

Published in:
Scandinavian Journal of Gastroenterology

DOI:
[10.3109/00365521.2015.1101486](https://doi.org/10.3109/00365521.2015.1101486)

2016

Document Version:
Peer reviewed version (aka post-print)

[Link to publication](#)

Citation for published version (APA):
Montén, C., Bjelkenkrantz, K., Gudjonsdottir, A. H., Browaldh, L., Arnell, H., Naluai, Å. T., & Agardh, D. (2016). Validity of histology for the diagnosis of paediatric coeliac disease: a Swedish multicentre study. *Scandinavian Journal of Gastroenterology*, 51(4), 427-433. <https://doi.org/10.3109/00365521.2015.1101486>

Total number of authors:
7

General rights

Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

**Validity of histology for the diagnosis of pediatric celiac disease:
a Swedish multicenter study**

Running headline: Histological diagnosis of celiac disease

Authors: Caroline Montén¹; Kaj Bjelkenkrantz²; Audur H. Gudjonsdottir³; Lars Browaldh⁴; Henrik Arnell⁵; Åsa Torinsson Naluai⁶; Daniel Agardh¹

Affiliations: Diabetes & Celiac Disease Unit, Department of Clinical Sciences, Lund University, Malmö, Sweden¹; Unilabs, Pathology, Sweden²; Queen Silvia Children's Hospital, Sahlgrenska Academy, Department of Pediatrics, Gothenburg, Sweden³; Department of Clinical Science & Education, Karolinska Institute Södersjukhuset, Stockholm, Sweden⁴; Department of Pediatric Gastroenterology, Hepatology & Nutrition, Karolinska University Hospital, Stockholm, Sweden⁵; Institute of Biomedicine, Department of Microbiology & Immunology, Sahlgrenska Academy, Gothenburg, Sweden⁶

Corresponding Author:

Daniel Agardh, Department of Clinical Sciences, Diabetes & Celiac Disease Unit, Lund University, Jan Waldenströms gata 35, CRC, 91:10, 202 05 Malmö, Sweden

E-mail: daniel.agardh@med.lu.se

Abstract

Objective: Histological evaluation of intestinal biopsies for the diagnosis of celiac disease can be challenging and compatible with risk for misdiagnosis. The aim was to evaluate the agreement of pathological diagnosis for celiac disease in children investigated at four major pediatric university hospitals in Sweden.

Material and Methods: Intestinal duodenal biopsies were collected from 402 children at median 9.7 years (1.4-18.3 years). A pathologist at each hospital performed the primary evaluation. A designated pathologist, blinded to the primary evaluation, performed a second Marsh classification of biopsies (M0 to M3c) taken from the bulb and duodenum separately. Kappa (κ) scores between first and second evaluation determined the agreement. Plasma samples were collected at the day of intestinal biopsy and analyzed for tissue transglutaminase autoantibodies (tTGA) using radio ligand binding assays.

Results: Marsh scores were concordant in 229/356 biopsies (64%, $\kappa=0.52$, $p<0.0001$). Among discordant results, 15/127 (12%) showed M0 in distal duodenum but \geq M2 in the bulb, whereas the opposite was true for 8/127 (6%) of the biopsies. There were fewer collected duodenal biopsies, more missing bulb biopsies and missing CD3 staining among discordant evaluations. The second evaluation revealed a Marsh score compliant with celiac disease in 22 children of whom 7 children were tTGA positive.

Conclusions: The variation between university hospitals on the pathological evaluation of biopsies may lead to misdiagnosis of celiac disease in pediatric patients. Access to clinical and endoscopic information as well as tTGA levels may be useful for the pathologist to complement the evaluation in dubious cases.

Keywords: Celiac disease, endoscopy, diagnosis, pediatrics, tissue transglutaminase autoantibodies.

Introduction

The histological evaluation of duodenal biopsies has been a corner stone for the diagnosis of celiac disease ever since the diagnostic criteria were first established in Interlaken 1969 [1]. The microscopic features showing villous atrophy, crypt hyperplasia and infiltration of intra-epithelial lymphocytes (IELs) that reverts to normal after treatment with gluten-free diet are typical appearances and exclusively seen in celiac disease [2]. With the discovery of disease specific autoantibodies and associated genetic markers over the past decades [3], the intestinal biopsy has been questioned as the current golden standard of methods for celiac disease diagnosis [4]. Recently, serology and HLA risk genotypes have been suggested to suffice in the diagnosis of cases with clear symptoms and high tissue transglutaminase autoantibody (tTGA) levels [5]. However, a substantial number of individuals detected with a positive tTGA serology may not have evident symptoms, clinical signs or a damaged intestinal mucosa [6, 7]. Therefore, a duodenal biopsy is still considered as an inevitable diagnostic tool to confirm or to exclude the diagnosis whenever a clinical suspicion of celiac disease is doubtful.

The histological classification according to Marsh is applied as a clinical routine worldwide and scores the severity of the histological lesion grading from a normal small intestinal mucosa to signs of inflammation featuring infiltration of IELs and deranged crypts and villous architecture [8]. However, with the number of histological categories, the inter-observer agreement for pathologists at biopsy classification decreases. The agreement is higher for biopsies showing normal features or a flat mucosa compared with those showing minimal signs of inflammation [9]. Evaluation of biopsies with minimal changes can therefore be challenging and compatible with risk for misdiagnosis. Factors such as an incorrectly oriented biopsy specimen, lack of staining IELs for CD3, missing biopsies from the duodenal bulb or insufficient number of biopsies collected from duodenum may increase the risk for

misclassification. Equally important, biopsy interpretations also distinguish between pathology practice settings and the experience of the observer, which may subsequently lead to misdiagnosis of the patient [10].

In this study, we reevaluated all biopsies collected from children included in a multicenter study of celiac disease, the Gene Expression (GENEX) study of Swedish children undergoing endoscopy. We aimed to test how the biopsy interpretation by pathologists, either at the same or at different pathology units, will have an impact on the diagnostic outcome of celiac disease in pediatric patients investigated with intestinal biopsy.

Methods

Study participants

Multiple biopsies were collected from the proximal duodenal bulb and distal duodenum of 402 children (243 females and 159 males) undergoing upper endoscopy for various reasons at median 9.7 years (1.4-18.3 years). All children were consecutively recruited between 2010 and 2012 at four pediatric departments situated in Malmö (Lab I), Gothenburg (Lab II), northern Stockholm (Lab III) and southern Stockholm (Lab IV). A total of 299 (74%) children were investigated at Lab I and 103 (26%) children investigated at Lab II-IV. Parents gave their informed written consent and the local ethical committee approved the study.

Endoscopic procedures & preparation of the biopsy specimen

Standard upper endoscopy with biopsies was performed under general anesthesia or deep sedation with Propofol. All biopsies were after sampling fixed in formalin, embedded in paraffin and thereafter cut serially in 4 µm sections and stained with hematoxylin and eosin according to standard procedure protocol. In addition, anti-CD3 monoclonal antibodies were at some hospitals used for immunostaining in order to simplify the IEL counting.

Primary classification of biopsies

All biopsies were first classified by an experienced local pathologist, unblinded to clinical history and macroscopic findings, at each of the four clinical centers as part of normal local clinical procedure. The biopsy showing the most severe finding, either located in the bulb or distal duodenum, determined the final Marsh score classified as: a Marsh score 0 (M0) was defined as a mucosa showing ≤ 25 IELs/100 enterocytes and normal villous and crypt architecture; M1 as M0 but with >25 IELs/100 enterocytes; M2 as M1 but with crypt hyperplasia; M3a as M2 but with partial villous atrophy; M3b as M3a but with subtotal

villous atrophy; M3c as M3b but with total villous atrophy. Lab I, Lab II and Lab IV applied the classification of Marsh [8], whereas Lab III applied the classification according to KVASt (KVAlitets och STandardiseringskommitté) score [11]. Lab IV did not distinguish between M3a, M3b and M3c classification, which were all merged to Marsh 3. KVASt scores biopsies according to four categories that do not distinguish M1 from M2 and consolidate subtotal and total villous atrophy in the same KVASt category. For the purposes of this study, KVASt was converted to a Marsh score. KVASt I was converted into M0-1, KVASt II into Marsh 2, KVASt III into M3a and KVASt IV into M3b-M3c. All comparisons within this study were conducted using four category Marsh score with M0 merged with M1 and M3b merged with M3c. Children with unspecified Marsh score at Lab IV were excluded from analysis. A biopsy showing at least a M2 was compatible with celiac disease.

Blinded re-evaluation of biopsies

After the first evaluation, a local pathologist in Malmö retrospectively performed a second blinded evaluation of all biopsies according to the Marsh classification, without any background information on clinical status, endoscopy findings or tTGA levels, or any knowledge about the initial diagnosis or outcome of the first evaluation. At the second evaluation, all biopsies were separately scored from the bulb and duodenum, respectively.

Marsh score agreement

Agreement on Marsh scores between the first and second evaluation as well as between the bulb and duodenum were determined with Kappa (κ) statistics [12]. A κ coefficient between 0.81 and 1.00 was considered a “very good agreement”; between 0.61 and 0.80 a “good

agreement”; between 0.41 and 0.60 a “moderate agreement”; between 0.21 and 0.40 a “fair agreement” and less than 0.20 a “poor agreement”.

Tissue transglutaminase autoantibodies (tTGA)

IgA anti-tTGA was measured separately using radioligand binding assay according to previously described methodology [13]. The cut-off for tTGA was set to 4 U/ml as estimated from ROC curves of healthy blood donors. Samples were analyzed in duplicates and antibody levels determined from a standard curve of pooled positive tTGA samples using the GraphPad PRISM 6.0 software.

Results

The average numbers of biopsies collected from the distal part of duodenum from each patient were six at Lab 1, five at Lab III and four at Lab II and Lab IV. In addition to distal duodenal biopsies, two separate biopsies on an average were collected from the bulb at each of the four clinics. Only 16% of bulb or duodenal biopsies were stained with anti-CD3 at Lab IV compared to 92% of the distal duodenal biopsies and 81% of bulb biopsies at Lab I, respectively. No biopsies were stained with anti-CD3 at Lab II or Lab III.

Marsh scores according to the first evaluation

Table 1 summarizes the first evaluation of biopsies from 381 children for each clinical center. Over all, biopsies from 217 (57%) children showed M0-M1, 14 (3.7%) M2, 49 (13%) M3a and 101 (27%) M3b-M3c. Lab IV classified biopsies from 21 children as M3 without specifying the degree of intestinal atrophy.

Marsh scores according to the second evaluation

Qualities of biopsies from four children were considered insufficient for an adequate second evaluation and another 21 children with unspecified Marsh at the first evaluation and were therefore excluded from further analysis. The second evaluation from the remaining 377 children resulted in 193 (51%) children to have M0-M1, 10 (2.7%) M2, 31 (8.2%) M3a and 143 (38%) M3b-M3c, respectively (Table 2).

The inter- and intra-variability of Marsh scores

Agreement on Marsh scores between the first and the second evaluation from of each clinical centers are given in Supplemental Tables 1-4. There was a good agreement between the first and blinded second evaluation at Lab 1, both between those performed by the same

pathologist (106/134, 79%, $\kappa=0.64$, $p<0.0001$) and between those performed by different pathologists (142/164, 87%, $\kappa=0.77$, $p<0.0001$). Furthermore, a moderate agreement was noted between the evaluations at Lab III (24/39, 62%, $\kappa=0.47$, $p<0.0001$) and a fair agreement between the evaluations at Lab II (9/13, 69%, $\kappa=0.38$, $p=0.025$) and Lab IV (19/27, 70%, $\kappa=0.36$, $p=0.002$).

The results of the first and second evaluation were discordant in 24/377 (6%) children, (Table 3). These discordant evaluations were 16/298 (5%) at Lab I, 2/13 (15%) at Lab II, 1/39 (3%) at Lab III and 5/27 (19%) at Lab IV. In four of the discordant evaluations, biopsies had not been collected from the duodenal bulb and in 18 of the patients the Marsh score differed for the duodenal bulb and distal duodenum. Two children were considered to have either crypt hyperplasia or villous atrophy at the first evaluation but not at the second evaluation. In the remaining 22 children with incoherent results, the second evaluation revealed a Marsh score consistent with celiac disease of whom 15 were tTGA negative and 7 were tTGA positive. No difference in scoring was noted for age, sex or other comorbidities between the two evaluations.

Marsh scores according to the location of the biopsy

Marsh scores after the second evaluation was further compared between biopsies collected either in proximal or in distal duodenum from the same patient. Concordant scorings were noted in 229/356 biopsies (64%, $\kappa=0.52$, $p<0.0001$). Among discordant results, 15/127 (12%) biopsies were classified as M0 from the duodenum and \geq M2 from the bulb, whereas the opposite situation was true for 8/127 (6%) of the biopsies (Table 4).

The inter- and intra-variability of Marsh scores with tTGA

The correlation was stronger after the first evaluation ($r=0.747$) (Fig. 1) than after the second evaluation ($r=0.634$) (Fig. 2), respectively ($p<0.001$). At the first evaluation, median tTGA levels were 1.1 U/ml for children with M0-M1, 42.3 U/ml with M2, 72.1 U/ml with M3a and 130.9 U/ml with M3b-M3c. At the second evaluation, median tTGA levels were 1.1 U/ml in children with M0-M1, 3.1 U/ml with M2, 3.8 U/ml with M3a and 98.0 U/ml with M3b and M3c. Among the 213 tTGA positive children, 77% of the biopsies demonstrated villous atrophy (\geq Marsh 3a) after the first evaluation which increased to 84% after the second evaluation (Table 5). When setting a tTGA threshold of 40 U/ml (i.e. the 10 fold value above the tTGA cutoff level of positivity), the proportion of children with villous atrophy were 91% after the first evaluation and 97% after the second evaluation, respectively. No apparent difference in proportion of children with villous atrophy between the first and the second evaluation was observed when increasing the tTGA cut-off level.

Discussion

The detection of histological alterations in the small intestinal mucosa is prone to inter-observer variability, which could result in patients being misdiagnosed [9, 10, 15]. In the present study, we show that the agreement of biopsy evaluation was higher for pathologists at the same department compared to that of between pathologists from different hospitals. Within one department, however, the same degree of agreement was observed between two different pathologists as between two evaluations by the same pathologist. These findings are in line with similar previous reports and stress the importance of local procedures on the outcome of the pathological diagnosis [10, 15]. Also in accordance with previous studies [9, 10], we found the lowest agreement in Marsh score between different clinical sites for biopsies showing minimal signs of inflammation.

In addition, we extended our biopsy findings with analysis of tTGA assessed in one laboratory using radioligand binding assays with proven high diagnostic performance [13]. We observed a good correlation between the degree of intestinal lesion and tTGA levels, indicating that tTGA is an important proxy for celiac disease and thus may support the pathologists in the final evaluation of the biopsy. However, despite the use of the same radioimmunoassay to determine tTGA levels for all the children in this study, the correlation between the degree of villous atrophy and tTGA levels was lower at the second evaluation compared to the first evaluation. Although various antibody tests with different cut-off levels were used to determine tTGA at the four clinical sites [16, 17], the status on tTGA levels probably affected the primary evaluation and thus resulted in a higher correlation between tTGA levels and degree of villous atrophy than in the second reevaluation where the pathologist was withheld this information.

It has been suggested that about 20-30% of celiac disease pediatric patients could be diagnosed based solely on high levels of tTGA without the need for an invasive endoscopic

procedure [3]. Current guidelines suggest that a tTGA level 10 times over the cut-off level of positivity could be optimal for omitting of duodenal biopsies in celiac disease diagnosis [14]. In our study, the proportion of children diagnosed with villous atrophy having tTGA levels 100 U/mL or above (the cut-off arbitrary defined as high level in our in-house radioligand binding assays), did not differ between the first and the second evaluation. All the discordant patients who would have received a change in diagnosis after the second evaluation had a tTGA level below 100 U/ml. In fact, a majority of these patients were given a Marsh score compliant with celiac disease despite being tTGA negative. Enteropathy or intraepithelial lymphocytosis can be present in other medical conditions than celiac disease [18], which obviously can complicate the biopsy evaluation procedure. Some of our discordant tTGA negative patients were diagnosed with Crohn's disease after the first evaluation. This suggest that risk of misclassification are mainly found among those with normal to milder forms of enteropathy and that a correct diagnosis of celiac disease in dubious cases will be difficult without any information on the tTGA status of a patient.

The abnormal histological features of the small intestine observed in celiac disease may show variable severity and could be patchy with signs of injury only in certain parts of the small intestine [19]. Therefore, during upper endoscopy it is recommended to collect at least one bulb biopsy and four biopsies from the more distal part of duodenum [14] and some clinicians are encouraged to avoid collecting bulb biopsies due to that the presence of Brunner's glands can complicate the pathological evaluation [20]. In this study, a present patchy mucosa as well as missing biopsies from the bulb seems to affect the diagnostic procedure negatively and lead to increased diagnostic discrepancies. Moreover, the proportion of discordant evaluations with changed diagnosis was higher for the two centers that had fewer biopsies collected from distal duodenum compared to the other centers. A majority of our patients given different Marsh scores in the proximal bulb and distal duodenum had more severe

findings in the bulb than distally, stressing the importance of collecting multiple biopsies as well as separate biopsies from the bulb.

In addition, CD3 staining for assessment of intraepithelial infiltration was missing at one of these two sites and only performed in a subset of samples at the other site. The importance of performing CD3 staining for improvement of the histological diagnosis of celiac disease was recently highlighted [21]. Still, an increased number of IELs may occur in other intestinal conditions as well and to use a fixed count as a cut-off normal may both cause interpretational difficulties [23]. It is therefore recommended that the scoring of IELs should be performed by an experienced pathologist and the interpretation should be made with caution by the physician carefully considering other intestinal disorders among cases with isolated M1.

Some limitations with this study need to be mentioned. Results on HLA and information on specific symptoms was not accessible. Recently, it was shown that a proper preceded clinical investigation of the patient showing clear symptoms of celiac disease strengthens the agreement between tTGA levels and the pathological biopsy evaluation [17]. Another study found that the correlation between tTGA levels and Marsh score were weaker in asymptomatic compared to symptomatic children with celiac disease [22], although patients within the latter study were detected by screening and may have been differently evaluated by the pathologists.

In summary, the reproducibility of a biopsy interpretation seems to be dependent on several determining factors for the final pathological diagnosis of celiac disease. Despite the fact that there are uniform criteria for the histological classification of celiac disease, evaluations still differ between pathologists and clinical sites, suggesting that local practices are an important factor for disparities between discordant results. In addition, the number of collected biopsies, the exact location in duodenum from where the biopsies are collected as well as if the biopsies are prepared with CD3 staining also seem to affect the pathological evaluation. This study

also indicates that the information on tTGA levels may be useful for the pathologist to complement the evaluation in dubious cases. In order to avoid misdiagnosis of celiac disease, education on scoring system and tTGA levels, minimal experience of the pathologist before evaluating intestinal biopies, minimal number of biopies from the duodenum and the bulb to ensure better assessments, cannot be overlooked for a correct diagnosis in pediatric patients in Sweden.

Acknowledgement

We would like to extend our gratitude to all the families and patients who contributed to the study.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

1. Mcneish AS, Harms HK, Rey J, Shmerling DH, Visakorpi JK, Walkersmith JA. Diagnosis of Celiac-Disease - Commentary on the Current Practices of Members of the European-Society-for-Paediatric-Gastroenterology-and-Nutrition (Espgan). Archives of disease in childhood. 1979;54(10):783-6.
2. Marsh MN. Gluten, Major Histocompatibility Complex, and the Small-Intestine - a Molecular and Immunobiological Approach to the Spectrum of Gluten Sensitivity (Celiac Sprue). Gastroenterology. 1992;102(1):330-54.
3. Giersiepen K, Lelgemann M, Stuhldreher N, Ronfani L, Husby S, Koletzko S, et al. Accuracy of Diagnostic Antibody Tests for Coeliac Disease in Children: Summary of an Evidence Report. Journal of pediatric gastroenterology and nutrition. 2012;54(2):229-41.
4. Walkersmith JA, Guandalini S, Schmitz J, Shmerling DH, Visakorpi JK. Revised Criteria for Diagnosis of Celiac-Disease. Archives of disease in childhood. 1990;65(8):909-11.
5. Ribes-Koninckx C, Mearin ML, Korponay-Szabo IR, Shamir R, Husby S, Ventura A, et al. Coeliac Disease Diagnosis: ESPGHAN 1990 Criteria or Need For a Change? Results of a Questionnaire. Journal of pediatric gastroenterology and nutrition. 2012;54(1):15-9.
6. Simell S, Kupila A, Hoppu S, Hekkala A, Simell T, Stahlberg MR, et al. Natural history of transglutaminase autoantibodies and mucosal changes in children carrying HLA-conferred celiac disease susceptibility. Scandinavian journal of gastroenterology. 2005;40(10):1182-91.
7. Kurppa K, Ashorn M, Iltanen S, Koskinen LLE, Saavalainen P, Koskinen O, et al. Celiac Disease without Villous Atrophy in Children: A Prospective Study. J Pediatr-US. 2010;157(3):373-U48.

8. Marsh et al. Mucosal histopathology in celiac disease: a rebuttal of Oberhuber's sub-division of Marsh III. *Gastroenterol Hepatol Bed Bench*. 2015 Spring;8(2):99-109.
9. Corazza GR, Villanacci V, Zambelli C, Milione M, Luinetti O, Vindigni C, et al. Comparison of the interobserver reproducibility with different histologic criteria used in celiac disease. *Clin Gastroenterol H*. 2007;5(7):838-43.
10. Arguelles-Grande C, Tennyson CA, Lewis SK, Green PHR, Bhagat G. Variability in small bowel histopathology reporting between different pathology practice settings: impact on the diagnosis of coeliac disease. *Journal of clinical pathology*. 2012;65(3):242-7.
11. Ludvigsson JF, Brandt L, Montgomery SM, Granath F, Ekbom A. Validation study of villous atrophy and small intestinal inflammation in Swedish biopsy registers. *Bmc Gastroenterol*. 2009;9.
12. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-74.
13. Kjelleras J, Vaziri-Sani F, Agardh D. Improved efficacy by using the pTnT-rhtTG plasmid for the detection of celiac disease specific tissue transglutaminase autoantibodies in radioligand binding assays. *Scand J Clin Lab Inv*. 2011;71(8):701-4.
14. Husby S, Koletzko S, Korponay-Szabo IR, Mearin ML, Phillips A, Shamir R, et al. European Society for Pediatric Gastroenterology, Hepatology, and Nutrition Guidelines for the Diagnosis of Coeliac Disease. *Journal of pediatric gastroenterology and nutrition*. 2012;54(1):136-60.
15. Weile B, Hansen BF, Hagerstrand I, Hansen JPH, Krasilnikoff PA. Interobserver variation in diagnosing coeliac disease. A joint study by Danish and Swedish pathologists. *Apmis*. 2000;108(5):380-4.

16. Egner W, Shrimpton A, Sargur R, Patel D, Swallow K. ESPGHAN Guidance on Coeliac Disease 2012: Multiples of ULN for Decision Making Do Not Harmonise Assay Performance Across Centres. *Journal of pediatric gastroenterology and nutrition*. 2012;55(6):733-5.
17. Vermeersch P, Geboes K, Marien G, Hoffman I, Hiele M, Bossuyt X. Defining Thresholds of Antibody Levels Improves Diagnosis of Celiac Disease. *Clin Gastroenterol H*. 2013;11(4):398-403.
18. Serra S, Jani PA. An approach to duodenal biopsies. *Journal of clinical pathology*. 2006;59(11):1133-50.
19. Ravelli A, Villanacci V, Monfredini C, Martinazzi S, Grassi V, Manenti S. How Patchy Is Patchy Villous Atrophy?: Distribution Pattern of Histological Lesions in the Duodenum of Children With Celiac Disease. *American Journal of Gastroenterology*. 2010;105(9):2103-10.
20. Vogelsang H, Hanel S, Steiner B, Oberhuber G. Diagnostic duodenal bulb biopsy in celiac disease. *Endoscopy*. 2001;33(4):336-40.
21. Mubarak A, Wolters VM, Houwen RH, Ten Kate FJ. Immunohistochemical CD3 staining detects additional patients with celiac disease. *World journal of gastroenterology : WJG*. 2015;21(24):7553-7.
22. Agardh D, Lee HS, Kurppa K, Simell V, Aronsson CA, Jorneus O, et al. Clinical features of celiac disease: a prospective birth cohort. *Pediatrics*. 2015;135(4):627-34.
23. Rostami K, Aldulaimi D, Holmes G, Johnson MW, Robert M, Srivastava A, et al. Microscopic Enteritis; Bucharest consensus. *WJG* 2015: 7, 2593-2604.

Table 1. Number of patients with Marsh score of biopsies at the first evaluation. Total N=381

| Marsh score/ center | M0-M1 | M2 | M3a | M3b-3c |
|--------------------------------|--------------|-----------|------------|---------------|
| Lab I | 167 | 5 | 37 | 90 |
| Lab II | 10 | 0 | 2 | 1 |
| Lab III | 16 | 6 | 10 | 8 |
| Lab IV | 24 | 3 | 0 | 2 |
| Total | 217 | 14 | 49 | 101 |

Table 2. Number of patients with Marsh score of biopsies at the blinded second evaluation.

Total N=377

| Marsh score/ center | M0-M1 | M2 | M3a | M3b-3c |
|--------------------------------|--------------|-----------|------------|---------------|
| Lab I | 154 | 7 | 21 | 116 |
| Lab II | 8 | 1 | 1 | 3 |
| Lab III | 14 | 0 | 7 | 18 |
| Lab IV | 17 | 2 | 2 | 6 |
| Total | 193 | 10 | 31 | 143 |

Table 3. Marsh scores, initial diagnosis and IgA-tTG levels of subjects with discordant results between the first and second (blinded) evaluation.

| Lab | First evaluation | | Second evaluation | | IgA-tTG level (U/ml) |
|---------|------------------|----------------------------------|-----------------------------|-------------------------|----------------------|
| | Marsh score | Initial diagnosis ¹ | Marsh score of the duodenum | Marsh score of the bulb | |
| Lab I | M0 | Crohn's disease | M2 | M3a | 0 |
| Lab I | M0 | Crohn's disease | M3a | NA* | 0 |
| Lab I | M0 | Crohn's disease | M0 | M3a | 3 |
| Lab I | M0 | H. pylori gastritis | M2 | M0 | 3 |
| Lab I | M0 | H. pylori gastritis | M2 | NA* | 2 |
| Lab I | M1 | Gastro-esophageal reflux disease | M1 | M3a | 2 |
| Lab I | M3a | Celiac disease | M1 | NA* | 47 |
| Lab I | M2 | Celiac disease | M1 | M1 | 15 |
| Lab I | M1 | Gastritis | M3a | M0 | 0 |
| Lab I | M1 | Gastro-esophageal reflux disease | M1 | M3b | 0 |
| Lab I | M1 | Gastritis | M1 | M3a | 1 |
| Lab I | M1 | No diagnosis | M2 | M1 | 4 |
| Lab I | M1 | No diagnosis | M3b | M3a | 26 |
| Lab I | M1 | Constipation | M2 | NA* | 1 |
| Lab I | M1 | No diagnosis | M1 | M3a | 80 |
| Lab I | M1 | Recurrent abdominal pain | M2 | M0 | 0 |
| Lab II | M0-M1 | Food intolerance | M0 | M3a | 2 |
| Lab II | M0-M1 | Gastro-esophageal reflux disease | M0 | M2 | 4 |
| Lab III | M0-M1 | T1 diabetes | M0 | M3a | 21 |
| Lab IV | M0-M1 | No diagnosis | M3a | M3a | 1 |
| Lab IV | M0-M1 | No diagnosis | M3a | M3b | 8 |
| Lab IV | M0-M1 | No diagnosis | M3a | M3b | 24 |
| Lab IV | M0-M1 | No diagnosis | M1 | M2 | 1 |
| Lab IV | M0-M1 | No diagnosis | M2 | M0 | 3 |

¹Diagnosis based solely on worse Marsh score

*NA, Not applicable.

Table 4. Number of patients with Marsh score of biopsies collected from either the bulb or distal duodenum.

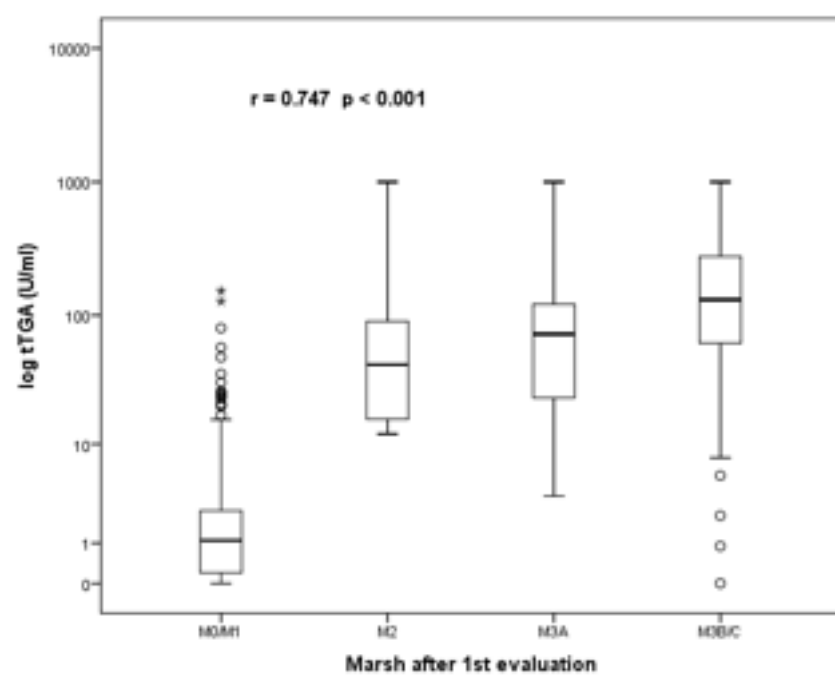
| Marsh score/ location | | Bulb biopsy | | | | | | Total |
|------------------------------|-----|-------------|----|----|-----|-----|-----|-------|
| | | M0 | M1 | M2 | M3a | M3b | M3c | |
| Distal duodenal biopsy | M0 | 120 | 5 | 1 | 8 | 4 | 2 | 140 |
| | M1 | 27 | 17 | 1 | 6 | 6 | 0 | 57 |
| | M2 | 4 | 1 | 1 | 2 | 3 | 0 | 11 |
| | M3a | 3 | 0 | 1 | 8 | 16 | 2 | 30 |
| | M3b | 1 | 1 | 0 | 9 | 29 | 17 | 57 |
| | M3c | 0 | 0 | 0 | 0 | 7 | 54 | 61 |
| Total | | 155 | 24 | 4 | 33 | 65 | 75 | 356 |

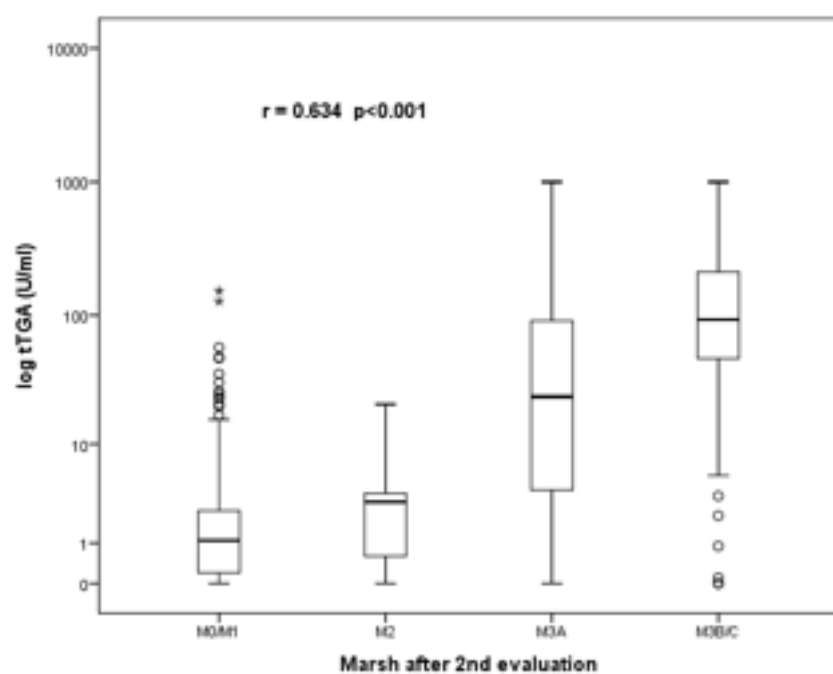
Table 5. Proportion of patients with villous atrophy (\geq Marsh 3a) according to different IgA-tTGA threshold values (U/ml) after the first and the second evaluation.

| IgA-tTGA threshold | Atrophic at first evaluation N (%) | Atrophic at second evaluation N (%) |
|-------------------------------|---|--|
| >4 (cut-off) | 163 (77) | 178 (84) |
| >40 (cut-off x 10) | 134 (91) | 142 (97) |
| >80 (cut-off x 20) | 93 (93) | 98 (98) |

Fig. 1. Correlation between Marsh score and IgA-tTGA levels at the first evaluation.

Fig. 2. Correlation between Marsh score and IgA-tTGA levels at the second evaluation.





Supplemental Table 1. Marsh score agreement between the first evaluation by pathologists at Lab I and the blinded second evaluation.

| Lab I | | Second evaluation | | | | Total |
|------------------|-------|-------------------|----|-----|-------|-------|
| | | M0-M1 | M2 | M3a | M3b-c | |
| First evaluation | M0-M1 | 152 | 5 | 7 | 2 | 166 |
| | M2 | 1 | 1 | 1 | 2 | 5 |
| | M3a | 1 | 1 | 9 | 26 | 37 |
| | M3b-c | 0 | 0 | 4 | 86 | 90 |
| Total | | 154 | 7 | 21 | 116 | 298 |

Supplemental Table 2. Marsh score agreement between the first evaluation by pathologists at Lab II and the blinded second evaluation.

| Lab II | | Second evaluation | | | | Total |
|------------------|-------|-------------------|----|-----|-------|-------|
| | | M0-M1 | M2 | M3a | M3b-c | |
| First evaluation | M0-M1 | 8 | 1 | 1 | 0 | 10 |
| | M2 | 0 | 0 | 0 | 0 | 0 |
| | M3a | 0 | 0 | 0 | 2 | 2 |
| | M3b-c | 0 | 0 | 0 | 1 | 1 |
| Total | | 8 | 1 | 1 | 3 | 13 |

Supplemental Table 3. Marsh score agreement between the first evaluation by pathologists at Lab III and the blinded second evaluation.

| Lab III | | Second evaluation | | | | Total |
|------------------|-------|-------------------|----|-----|-------|-------|
| | | M0-M1 | M2 | M3a | M3b-c | |
| First evaluation | M0-M1 | 14 | 0 | 1 | 0 | 15 |
| | M2 | 0 | 0 | 4 | 2 | 6 |
| | M3a | 0 | 0 | 2 | 8 | 10 |
| | M3b-c | 0 | 0 | 0 | 8 | 8 |
| Total | | 14 | 0 | 7 | 18 | 39 |

Supplemental Table 4. Marsh score agreement between the first evaluation by pathologists at Lab IV and the blinded second evaluation.

| Lab IV | | Second evaluation | | | | Total |
|------------------|-------|-------------------|----|-----|-------|-------|
| | | M0-M1 | M2 | M3a | M3b-c | |
| First evaluation | M0-M1 | 17 | 2 | 1 | 2 | 22 |
| | M2 | 0 | 0 | 1 | 2 | 3 |
| | M3a | 0 | 0 | 0 | 0 | 0 |
| | M3b-c | 0 | 0 | 0 | 2 | 2 |
| Total | | 17 | 2 | 2 | 6 | 27 |