ORIGINAL RESEARCH

Validation of celiac disease diagnoses recorded in the Danish National Patient Register using duodenal biopsies, celiac disease-specific antibodies, and human leukocyte-antigen genotypes

Stine Dydensborg Sander[1–3]
Ketil Størdal[4,5]
Tine Plato Hansen[6]
Anne-Marie Nybo Andersen[7]
Joseph A Murray[8]
Søren Thue Lillevang[9]
Steffen Husby[1,2]

[1]Hans Christian Andersen Children's Hospital, Odense University Hospital; [2]Institute of Clinical Research, University of Southern Denmark; [3]Odense Patient Data Explorative Network (OPEN), Odense University Hospital, Odense, Denmark; [4]Mental and Physical Health, Norwegian Institute of Public Health, Oslo; [5]Department of Pediatrics, Ostfold Hospital Trust, Fredrikstad, Norway; [6]Department of Pathology, Hvidovre Hospital, [7]Department of Public Health, University of Copenhagen, Copenhagen, Denmark; [8]Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN, USA; [9]Department of Clinical Immunology, Odense University Hospital, Odense, Denmark

Correspondence: Steffen Husby
Hans Christian Andersen Children's Hospital, Odense University Hospital, 23C Kløvervænget, Odense C 5000, Denmark
Tel +45 6541 2090
Fax +45 6591 1862
Email steffen.husby@rsyd.dk

Purpose: The purpose of this study was to validate the celiac disease diagnoses recorded in the Danish National Patient Register. To validate the diagnoses, we used information on duodenal biopsies from a national register of pathology reports (the Patobank) and information on celiac disease-specific antibodies and human leukocyte antigen (HLA) genotypes obtained from patient medical records.

Patients and methods: We included all the children who were born from 1995 to 2012 and who were registered as having celiac disease in the Danish National Patient Register. We reviewed all the pathology reports on duodenal biopsies in the Patobank and the information in the medical records on celiac disease-specific antibodies (ie, anti-tissue transglutaminase 2 IgA and IgG, endomysial antibodies IgA, and anti-deamidated gliadin peptide IgG) and HLA genotypes.

Results: We identified 2,247 children who were registered in the Danish National Patient Register with celiac disease. Duodenal biopsies for 1,555 of the children (69%) were registered in the Patobank; 1,127 (50%) had a biopsy that was compatible with celiac disease (ie, Marsh 2–3). We accessed the medical records of 95% of the children who were registered in the Danish National Patient Register with celiac disease. We found that 1,510 (67%) had one or more positive antibody-test results; 1,120 (50%) had anti-tissue transglutaminase 2, IgA at tenfold or greater the upper limit of the normal range and/or positive endomysial antibody results. The positive predictive value depended on the criteria used for validation and the types and numbers of registrations that were included in the analysis and ranged from 62% (95% confidence interval: 60%–64%) to 86% (95% confidence interval: 84%–87%).

Conclusion: Our findings indicate that the Danish National Patient Register is a valuable source to identify patients who have been diagnosed with celiac disease. However, validation of the diagnoses is warranted before data on the patients are used for research purposes.

Keywords: administrative health register, national patient register, pathology register, medical record, histology, serology

Introduction

Celiac disease is a chronic immunomediated disorder that is elicited in genetically susceptible individuals by gluten and related prolamins. It is characterized by various combinations of clinical manifestations, celiac disease-specific antibodies, human leukocyte antigens (HLAs) DQ2 or DQ8, and enteropathy.

Celiac disease-specific antibodies, ie, anti-tissue transglutaminase 2 (tTG2) IgA or IgG, endomysial antibody (EMA) IgA, and anti-deamidated gliadin peptide (DGP) IgG, are reliable diagnostic tools. tTG2 IgA and EMA have sensitivities and specificities of ≥90%,[1] and they have been used in clinical practice in Denmark since the late 1990s. The HLA genotypes DQ2 or DQ8 are involved in the pathogenesis of celiac disease, and carriage of at least one of them is required for an individual to develop celiac disease. However, 30%–40% of the general population has HLA DQ2/DQ8. Therefore, testing for HLA DQ2/DQ8 is primarily used to exclude a diagnosis of celiac disease.[2] Celiac disease-related enteropathy is diagnosed by the histological evaluation of duodenal biopsies. The Marsh classification system can be used to classify the histopathological changes in duodenal biopsies, which are characteristic of but not specific to celiac disease. Enteropathy exists on a continuum of severity from intraepithelial lymphocytosis (Marsh 1), to hypertrophy of crypts (Marsh 2), to increasing villous atrophy (Marsh 3) and chronic inflammation of the lamina propria.[3]

The Danish national guidelines for the diagnosis of pediatric celiac disease are based on guidelines from the European Society for Paediatric Gastroenterology, Hepatology, and Nutrition (ESPGHAN). Until 2012, for a diagnosis to be made, the guidelines required a duodenal biopsy with histological changes that were compatible with celiac disease (Marsh 2–3).[4] New ESPGHAN guidelines, which were published in January 2012, allow for diagnoses of children with symptoms of celiac disease based on the presence of celiac disease-specific antibodies (ie, tTG2 IgA at ten or more times the upper limit of normal [ULN] and positive EMA results) and the presence of HLA DQ2/DQ8, omitting the biopsy.[2] These guidelines were implemented in Denmark in January 2013.

In Denmark, children are referred to a pediatric hospital department in order to obtain diagnoses of and treatment for celiac disease. All hospital visits are recorded in the Danish National Patient Register, which is a mandatory administrative health register that is considered to be one of the most comprehensive patient registers in the world.[5] The Danish National Patient Register has been used for reimbursing hospitals since 2000.[6] Therefore, the completeness of the records on celiac disease diagnoses in the register is expected to be high. However, the validity of the records is unclear. We previously demonstrated that from 1995 to 2010, approximately two in three children who were registered as having celiac disease in the Danish National Patient Register had biopsies that were compatible with celiac disease registered in the Patobank, a national Danish database of pathology reports.[7]

Numerous methods have previously been used to identify and validate the celiac disease diagnoses in various types of registers.[8–11] Incidence registers[12–15] and reimbursement registers[16] tend to use strict diagnostic criteria, which means that they often have high levels of validity, though their availability tends to be limited. Pathology reports on biopsies have previously been used to confirm registered diagnoses of celiac disease,[7,10,17] but the implementation of the new diagnostic guidelines for celiac disease means that this option is now inadequate.

The objective of this study was to validate the records of celiac disease in the Danish National Patient Register for children and adolescents in Denmark who were born from 1995 to 2012. We used a combination of pathology reports on duodenal biopsies that were registered in the Patobank and data on both celiac disease-specific antibodies and HLA genotypes that were obtained from patient medical records.

Patients and methods

We included all the children born from January 1, 1995 to December 31, 2012 who had been registered as having celiac disease in the Danish National Patient Register by May 8, 2015. We combined this information with the records of duodenal biopsies in the Patobank and with data on celiac disease-specific antibodies and HLA genotypes that were obtained from medical records. To mitigate against missing test results in the medical records, we also obtained data on celiac disease-specific antibodies and HLA genotypes from the laboratory at the Statens Serum Institut (SSI) and the GastroLab at Medical Laboratory F, Gentofte Hospital in Denmark.

To evaluate the completeness of celiac disease registrations in the Danish National Patient Register, we included children who were born from 1995 to 2012, were not registered as having celiac disease in the Danish National Patient Register, and had a duodenal biopsy recorded in the Patobank that was compatible with Marsh 1 or Marsh 2–3, and/or a record of a positive test result for celiac disease-specific antibodies at the SSI or the GastroLab.

A unique ten-digit personal registration number is assigned to everyone who is born or lives in Denmark. All public records are linked to these numbers, and they enable the individual-level linkage of information from various registers and medical records.[18]

The Danish National Patient Register

The Danish National Patient Register includes information on the dates of hospital admissions, discharges, and outpatient visits, and the associated hospitals, departments, and

diagnoses. From 1994 onward, the diagnoses recorded in the register have been based on the *International Classification of Diseases* (ICD)-10. Each record involves one primary diagnosis, and if relevant one or more secondary diagnoses or supplementary codes. The diagnoses can be tentative (ICD-10 Z03*).[5,6] We included all children who were born from 1995 to 2012 and who were registered as having a diagnosis of celiac disease (ICD-10 K900).

The Patobank

The Patobank is a national Danish pathology database that includes pathology reports containing information on the dates of duodenal biopsies, macroscopic and microscopic descriptions of the biopsies, the conclusions, and topographic, morphologic, and diagnostic codes based on the Systematized Nomenclature of Medicine (SNOMED; http://www.patobank.dk). We applied for all the records up to 2015 for children who were born from 1995 to 2012 and who were registered in the Danish National Patient Register as having a diagnosis of celiac disease.

Furthermore, to identify the children who were diagnosed with celiac disease but not registered in the Danish National Patient Register, we applied for all the records that had topographic, morphologic, and diagnostic codes that were relevant to celiac disease (Table S1) for children who were born from 1995 to 2012, regardless of their registration status in the Danish National Patient Register. These data were available only for 1995–2012.

All the pathology reports on duodenal biopsies were reviewed by one of the authors (SDS), and were classified as being compatible with Marsh 2–3 (ie, intraepithelial lymphocytosis, hypertrophy of crypts, and increasing villous atrophy), Marsh 1 (ie, intraepithelial lymphocytosis), Marsh 0 (ie, normal or no changes related to celiac disease), or as being unclassifiable.[3] In addition, a total of 100 randomly selected pathology reports were reviewed by another author (TPH). The two reviewers agreed on each of the 100 classifications.

Medical records

In Denmark, data on celiac disease-specific antibodies and HLA DQ2/DQ8 are recorded as laboratory-test results in patient medical records. To access relevant information for register-based research from medical records, permission is required from the Danish health authorities and subsequently from the hospital departments responsible for the treatment of the patients in question.

We aimed to access the medical records of all the children who were born from 1995 to 2012 and who were registered in the Danish National Patient Register as having celiac disease.

For each patient, we assumed that the department that was responsible for celiac disease-related treatments was the pediatric or internal medicine department that was associated with the most recent registration of celiac disease. If all the registrations of celiac disease for a patient were associated with departments that would not be expected to treat children with celiac disease (eg, orthopedic surgery departments), we contacted the department that was associated with the most recent registration of celiac disease.

After we had been granted permission, the medical records were reviewed either electronically, in person at the departments, or by requesting that photocopies of the records be mailed by the departmental staff. We extracted data on all the test results and their ULN associated with tTG2 IgA and IgG, EMA, and DGP antibodies, as well as the HLA DQ2/DQ8 test results. Information on the manufacturers of the test kits was not available.

Additional laboratory data

In Denmark, computerization of medical records began during the study period 1995–2012. Computerized test results are less likely to be missed when reviewing medical records. Therefore, we expected that there would be more missing test results for records that were registered early in the study period. During the study period, tests for celiac disease-specific antibodies and HLA genotypes were analyzed by a small number of laboratories, primarily the SSI laboratory and the GastroLab. Therefore, we included all the test results from these two laboratories to increase the completeness of data. We were able to obtain 2000–2015 data from the SSI and 1997–2009 data from the GastroLab. These data included information about positive test results for children who were born from 1995 to 2012 but who were not registered as having celiac disease in the Danish National Patient Register.

Validity and completeness of celiac disease diagnoses in the Danish National Patient Register

In accordance with the criteria in Table 1, we validated the celiac disease diagnoses recorded in the Danish National Patient Register by combining data on biopsy classifications and data on celiac disease-specific antibodies and HLA genotypes. To estimate the completeness of the registrations of celiac disease in the Danish National Patient Register, we identified the children who did not have registrations of celiac disease in the Danish National Patient Register but who were registered in the Patobank with biopsies that were compatible with Marsh 1–3 or who had positive test results

Table 1 Assessments of celiac disease diagnoses

Assessment[a]	Biopsy classification	Celiac disease-specific antibodies	Additional criteria	Registration status in the Danish National Patient Register, n (%)	
				Registered	Not registered
Definite	Compatible with Marsh 2–3	Positive		969 (43)	9 (4)
	Any classification	tTG2 IgA at ≥10× ULN and positive EMA	HLA-positive	78 (3)	0
	Any classification	tTG2 IgA at ≥10× ULN or positive EMA	Recorded after January 1, 2013	176 (8)	5 (2)
	Compatible with Marsh 2–3	Only negative or no records		158 (7)	49 (20)
	Compatible with Marsh 1	tTG2 IgA at ≥10× ULN or positive EMA		22 (1)	1 (<1)
Probable	Unclassifiable or compatible with Marsh 0 or no biopsy recorded	tTG2 IgA at ≥10× ULN or positive EMA		95 (4)	22 (9)
Possible	Compatible with Marsh 1 or unclassifiable	Positive		156 (7)	69 (28)
	Compatible with Marsh 0 or no biopsy recorded				
Unclear (insufficient data)	Compatible with Marsh 1 or unclassifiable or compatible with Marsh 0 or no biopsy recorded	Unknown	Medical records not accessed	65 (3)	88 (36)
Unlikely	Compatible with Marsh 1 or unclassifiable or compatible with Marsh 0 or no biopsy recorded	Only negative or no records	Medical records accessed	492 (22)	0
	Compatible with Marsh 1 or unclassifiable or compatible with Marsh 0 or no biopsy recorded	Any antibody records	HLA-negative	36 (2)	1 (<1)

Note: [a]The authors' assessments of the likelihoods of the diagnoses of celiac disease being correct were based on the biopsy classifications, celiac disease-specific antibodies, and HLA genotypes.
Abbreviations: ULN, upper limit of normal; EMA, endomysial antibody; HLA, human leukocyte antigen; tTG2, anti-tissue transglutaminase 2.

for celiac disease-specific antibodies according to the data from the SSI laboratory or the GastroLab.

Statistical analysis

Positive predictive values (PPVs) were calculated as the proportion of children that fulfilled the validation criteria devised by the authors (Table 1) out of all the children who were registered with celiac disease in the Danish National Patient Register. We used the Wilson method to calculate 95% confidence intervals (CIs) in Stata 14 (StataCorp LP, College Station, TX, USA).

Ethics

The study was approved by the Danish Data Protection Agency and the Danish Health Authorities, and for cases where the medical records were accessed by the departments responsible for treatment of the patient at the following hospitals: Viborg Regional Hospital, Silkeborg Regional Hospital, Rigshospitalet, Herlev Hospital, Frederiksberg Hospital, Gentofte Hospital, Hvidovre Hospital, Nordsjællands Hospital, Holbæk Hospital, Nykøbing F. Hospital, Næstved Hospital, Slagelse Hospital, Zealand University Hospital Køge, Zealand University Hospital Roskilde, Odense University Hospital, Hospital Little Belt Kolding, Hospital Little Belt Vejle, South-West Jutland Hospital, South Jutland Hospital, Aarhus University Hospital, Horsens Regional Hospital, Randers Regional Hospital, Aalborg University Hospital Aalborg, Aalborg University Hospital Hobro, Herning Regional Hospital, North Denmark Regional Hospital Hjørring, and North Denmark Regional Hospital Thisted. Patient consent from the participants and notification of the regional committees on health-researchs ethics was not required as this was a register-based study not including human biological material, and the participants were not contacted (Health legislation §46.2 and Legislation for Health-Research-Ethics and Health Science Research §14.2).

Results

The Danish National Patient Register

We identified 2,247 children registered with a primary, secondary, or tentative diagnosis of celiac disease in the Danish National Patient Register, which corresponded to a total of 20,776 hospital visits. The first registration was on February 8, 1996, and the most recent registration was on May 7, 2015. The median number of hospital visits was six (range 1–143). Inpatient admissions accounted for 1,459 of the hospital visits. With regard to tentative diagnoses, 147 (6.5%) of the children had solely tentative diagnoses. With regard to the number of registrations of celiac disease, 322 (14.3%) of the children had one registration, and 1,778 (79.1%) had two or more registrations (Table 2).

The Patobank

We received all registrations of duodenal biopsy procedures for 1,555 of the children in the Danish National Patient Register (Table 3). Each registration was unique in terms of the patient and date of the procedure, and each registration could include more than one duodenal biopsy (taken during the same procedure). We identified 1,170 registrations (involving 1,127 children) that were compatible with Marsh 2–3, 126 registrations (involving 101 children) that were compatible with Marsh 1, 461 registrations (involving 298 children) that were compatible with Marsh 0, and 33 registrations (involving 29 children) that were unclassifiable. The unclassifiable biopsies were mostly of very poor quality, and so they could not be evaluated by a pathologist.

Among the children with Patobank records that included celiac disease-related SNOMED codes (Table S1) but who were not registered in the Danish National Patient Register, we identified 63 registrations (involving 58 children) that were compatible with Marsh 2–3 (Table 4), 94 registrations (involving 90 children) that were compatible with Marsh 1 (Table 4), and 628 registrations that were compatible with

Table 2 Assessment of the likelihood of celiac disease.

Assessment of celiac disease likelihood[a]	All registrations (n=2,247), n (%)	Excluding tentative diagnoses (n=2,100), n (%)	Two or more registrations (n=1,778), n (%)	No registrations[b] (n=244), n (%)	Total (n=2,491), n (%)
Definite	1,403 (62)	1,383 (66)	1,309 (74)	64 (26)	1,467 (59)
Probable	95 (4)	91 (4)	84 (5)	22 (9)	117 (5)
Possible	156 (7)	138 (7)	97 (5)	69 (28)	225 (9)
Unclear	65 (3)	57 (3)	33 (2)	88 (36)	153 (6)
Unlikely	528 (24)	431 (21)	255 (14)	1 (<1)	529 (21)

Notes: [a]See criteria in Table 1; [b]not registered in the Danish National Patient Register (these patients had biopsies that were compatible with celiac disease or positive antibody-test results registered with the Statens Serum Institut laboratory or the GastroLab).

Table 3 Registrations in the Patobank and antibody-test results for children registered with celiac disease in the Danish National Patient Register

Antibody-test results from medical records	Registration in the Patobank					Total
	Yes	Marsh 2–3	Marsh 1	Unclassifiable or Marsh 0	No	
Yes	1,484 (+24[a])	1,084 (+10[a])	92 (+4[a])	308 (+10[a])	651 (+6[a])	2,135 (+30[a])
One or more registrations	1,450	1,070 (+10[a])	89 (+4[a])	291 (+10[a])	583 (+6[a])	2,033 (+30[a])
One or more positive registrations	1,182	963 (+10[a])	75 (+10[a])	144 (+10[a])	319 (+6[a])	1,501 (+30[a])
Positive EMA or tTG2 IgA at ≥10× ULN	861	744 (+5[a])	35	82	253 (+1[a])	1,114 (+6[a])
No	71	43	9	19	41	112
Total	1,555	1,127	101	327	692	2,247

Note: [a]Medical records not accessed; antibody-test results were obtained from the Statens Serum Institut laboratory or the GastroLab.
Abbreviations: EMA, endomysial antibody; tTG2, anti-tissue transglutaminase 2; ULN, upper limit of normal.

Table 4 Registrations in the Patobank and antibody-test results for children not registered with celiac disease in the Danish National Patient Register

Antibodies from the SSI or the GastroLab	Registration in the Patobank				Total
	Yes	Marsh 2–3	Marsh 1	No	
Yes	30	18	12	96	126
Positive EMA results or tTG2 IgA at ≥10× ULN	8	7	1	27	35
No	118	40	78	NR	118
Total	148	58	90	96	244

Abbreviations: EMA, endomysial antibody; NR, not relevant; SSI, Statens Serum Institut; tTG2, anti-tissue transglutaminase 2; ULN, upper limit of normal.

Marsh 0 or that were unclassifiable. Table S2 shows the frequencies of selected SNOMED codes and the Marsh classifications of the biopsies.

Medical records

We contacted 72 hospital departments to obtain permission to access the medical records of the 2,247 children who were registered in the Danish National Patient Register. In total, we obtained access to medical records for 2,135 (95%) of the children (Table S3). Pediatrics (ie, all 18 of the pediatric departments in Denmark) accounted for 1,936 of the children (median: 100.5, range: 11–276), and 29 departments of internal medicine accounted for 245 children (median: 6, range: 1–42). Furthermore, three departments (which accounted for four children) were in private hospitals, ten were surgical departments (which accounted for 48 children), three were emergency departments (which accounted for three children), and nine were other types of departments (which accounted

for eleven children). In the medical records, we identified data on 11,204 blood samples that were unique in terms of the patient and the date of the blood collection that had tested for or IgG, EMA, or DGP antibodies or HLA DQ2/DQ8 (Tables 3 and 5).

Additional laboratory data

In the data from the SSI laboratory and the GastroLab, we found 48 additional records (involving 30 children) of celiac disease-specific antibodies (Table 3) for children who were registered as having celiac disease in the Danish National Patient Register but whose medical records (associated with 12 departments) were not accessed (see Table S3 for details). Furthermore, for children who were not registered as having celiac disease in the Danish National Patient Register, we found 167 positive test results (involving 126 children), including 94 positive results for tTG2 IgA at ten or more times the ULN or for EMA (involving 80 children) (Table 4).

Table 5 Assessments of the likelihood of celiac disease for various registrations in the Danish National Patient Register

Characteristics	Total (n=2,247)	Assessment[a]				
		Definite (n=1,403)	Probable (n=95)	Possible (n=156)	Unclear (n=65)	Unlikely (n=528)
Age at the end of the study period (May 8, 2015), years[b]	13.1 (9.3–16.7)	13.2 (9.7–17.7)	14.7 (10.5–16.9)	11.6 (8–15.9)	15.8 (11.4–18.8)	12.7 (8.9–16.3)
Female[c]	1,362 (61)	902 (64)	61 (64)	98 (63)	41 (63)	260 (49)
Danish National Patient Register						
Age at first registration, years[b]	4.2 (1.9–7.6)	4 (1.8–7.3)	4.7 (2.9–9.9)	2.6 (1.2–4.7)	2.9 (1.1–6.5)	5.7 (2.6–9.4)
Registrations per child[b]	6 (2–12)	8 (4–15)	8 (4–13)	3 (1–6)	2 (1–4)	2 (1–3.5)
1 registration[c]	417 (19)	87 (6)	10 (11)	50 (32)	30 (46)	240 (45)
2–4 registrations[c]	565 (25)	287 (20)	17 (18)	51 (33)	21 (32)	189 (36)
5–9 registrations[c]	551 (25)	418 (30)	34 (36)	30 (19)	6 (9)	63 (12)
10–14 registrations[c]	307 (14)	258 (18)	13 (14)	15 (16)	4 (6)	17 (3)
≥15 registrations[c]	407 (18)	353 (25)	21 (22)	10 (6)	4 (6)	19 (4)
Patobank						
Children with biopsy registration[c]	1,555 (69)	1,188 (85)	57 (60)	93 (60)	25 (39)	192 (36)
Age at first registration, years[b]	5.1 (2.5–8.1)	5 (2.5–7.7)	6 (4.2–10.6)	3.1 (1.9–5.2)	3.1 (2.1–6.3)	7.1 (4.2–10.4)
Antibody tests						
Antibody tests per child (dates)[b]	4 (2–7)	6 (4–8)	7 (5–9)	3 (2–6)	0	1 (1–3)
Age at first test, years[b]	4.7 (2.3–7.8)	4.5 (2.2–7.3)	5.7 (3.3–9.6)	2.9 (1.6–6)	10.4 (6.8–12.6)	6.2 (3.1–9.5)
Age at first positive test, years[b]	3.6 (1.9–6.7)	3.8 (1.9–6.9)	5.5 (3.2–9)	2.5 (1.3–4.2)	–	2.7 (1.4–4.2)
tTG2 IgA tests per child[b]	7 (4–10)	7 (5–10)	8 (6–11)	5 (3–7)	0 (0–1)	3 (1–5)
Children with tTG2 IgA tests[c]	2,043 (91)	1,351 (96)	95 (100)	156 (100)	14 (22)	427 (81)
Children with positive tTG2 IgA tests[c]	1,443 (64)	1,208 (86)	94 (99)	131 (84)	0	10 (2)
Children with tTG2 IgA at ≥10× ULN[c]	1,011 (45)	938 (67)	73 (77)	0	0	0
Children with EMA tests[c]	688 (31)	471 (34)	41 (43)	37 (24)	4 (6.2)	135 (26)
Children with positive EMA tests[c]	429 (19)	392 (28)	37 (39)	0	0	0
Children with tTG2 IgG tests[c]	478 (21)	304 (22)	25 (26)	35 (22)	1 (2)	113 (21)
Children with positive tTG2 IgG tests[c]	145 (7)	121 (9)	9 (10)	15 (10)	0	0
Children with DGP tests[c]	964 (43)	701 (50)	31 (33)	105 (67)	1 (1.5)	126 (24)
Children with positive DGP tests[c]	447 (20)	360 (26)	19 (20)	63 (40)	0	5 (0.95)
HLA genotype						
Children with HLA tests[c]	420 (19)	255 (18)	21 (22)	52 (33)	3 (4.6)	89 (17)
Children with HLA DQ2/DQ8[c]	379 (17)	251 (18)	20 (21)	52 (33)	3 (4.6)	53 (10)

Notes: [a]See criteria in Table 1; [b]median (interquartile range); [c]number (%).
Abbreviations: ULN, upper limit of normal; EMA, endomysial antibody; DGP, deamidated gliadin peptide; HLA, human leukocyte antigen.

Validity of celiac disease diagnoses in the Danish National Patient Register

Table 2 shows the distribution of celiac disease registrations in the Danish National Patient Register according to the criteria in Table 1 for definite, probable, possible, unclear, and unlikely diagnoses of celiac disease. Of the 2,247 children registered as having celiac disease in the Danish National Patient Register, 1,403 were considered to be correctly registered with celiac disease (ie, they had definite diagnoses). Another 95 children probably had celiac disease, based on the fact that they had tTG2 IgA at ten or more times the ULN or positive EMA results, but lacked evidence of histological changes. In addition, 156 children possibly had celiac disease, based on the fact that their test results were positive for celiac disease-specific antibodies, but they lacked evidence of histological changes. However, 65 cases were unclear due to insufficient information (ie, we could not access the relevant medical records), and 528 children were unlikely to have been correctly registered in the Danish National Patient Register. These results correspond to a PPV that ranged from 62.4% (95% CI: 60.4%–64.4%) if only the definite diagnoses are taken into account to 76.5% (95% CI: 74.7%–78.2%) if the definite, probable, possible, and unclear diagnoses are taken into account.

If we exclude the children who had only tentative diagnoses registered in the Danish National Patient Register, the number of children who were unlikely to have been correctly registered in the Danish National Patient Register decreases to 431. This results in a PPV that ranges from 65.9% (95% CI: 63.8%–67.9%) if only the definite diagnoses are taken into account to 79.4% (95% CI: 77.7%–81.1%) if the definite, probable, possible, and unclear diagnoses are taken into account.

If we include only children with two or more registrations of celiac disease in the Danish National Patient Register, the number of children who were unlikely to have been correctly registered in the Danish National Patient Register decreases to 255. This results in a PPV that ranges from 73.6% (95% CI: 71.5%–75.6%) if only the definite diagnoses are taken into account to 85.6% (95% CI: 84%–87.2%) if the definite, probable, possible, and unclear diagnoses are taken into account.

As indicated in Figure 1, the PPVs increased over time during the study period. The proportion of children who were registered with celiac disease in the Danish National Patient Register and who also had a duodenal biopsy registered in the Patobank dropped from 78% for those registered in 1996–2012 to 56% for those registered in 2013–2015 (data not shown). The proportion of biopsies that were compatible with Marsh 2–3 remained the same over time (73% for
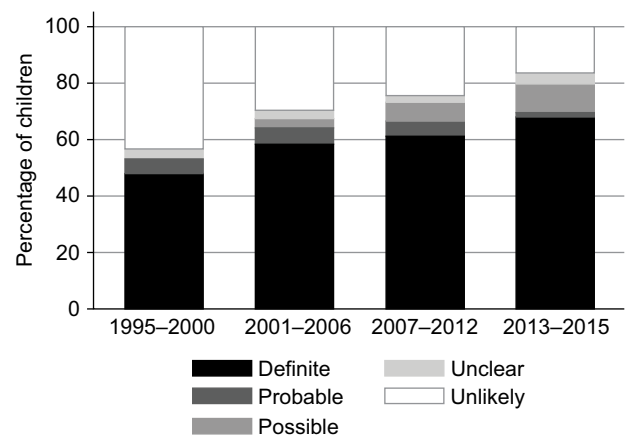

Figure I Assessment of celiac disease by year of first registration in the Danish National Patient Register.

those registered in 1996–2012 and 70% for those registered in 2013–2015; data not shown).

Completeness of celiac disease diagnoses in the Danish National Patient Register

We categorized 20 children who were recorded in the Danish National Patient Register as having only tentative diagnoses of celiac disease as having definite diagnoses of celiac disease. We identified an additional 64 children who were not recorded as having celiac disease in the Danish National Patient Register, but who fulfilled the criteria for a definite diagnosis of celiac disease. These children were either registered in the Patobank with a biopsy that was compatible with Marsh 2–3 or records from the SSI laboratory or the GastroLab showed that they had tTG2 IgA at ten or more times the ULN or positive EMA results (Table 2).

Discussion

We carried out a national validation study of celiac disease diagnoses that was based on duodenal biopsies, celiac disease-specific antibodies, and HLA genotypes. We found that the PPV of celiac disease diagnoses depended on the criteria used for validation and the types and numbers of registrations in the Danish National Patient Register that were included in the analysis. The PPV ranged from 62% (95% CI: 60%–64%) if only the definite diagnoses were taken into account to 86% (95% CI: 84%–87%) if only patients with two or more registrations were included and the definite, probable, possible, and unclear diagnoses were taken into account. To our knowledge, this is the first study to include information on biopsies, antibodies, and HLA genotypes to validate celiac disease diagnoses in a national patient register. Our findings are in-line with results from other validation studies of celiac disease diagnoses.

In the Norwegian National Patient Register, 82%–83% of the children who were registered from 2008 to 2009 had a second diagnosis registered from 2008 to 2011, which increased the likelihood of having a correct diagnosis.[8] The diagnoses of children that participated in the Norwegian Mother and Child Cohort Study were validated by collecting data using postal questionnaires sent to the families of children with two or more registrations of celiac disease in the Norwegian National Patient Register (n=468). The response rate was 63%, and 94% of the responses confirmed the diagnoses (83% of cases involved biopsy confirmations).[9]

In an Italian study in the region of the Friuli-Venezia Giulia region, 71% of children with a hospital-discharge diagnosis of celiac disease (n=1,227) had a relevant code recorded in a pathology register.[10]

In the UK, a single celiac disease registration per patient with the Clinical Practice Research Datalink service had a PPV of 81%, which increased to 89% after including only patients who had received gluten-free food prescriptions and to 100% after including only patients who had two diagnostic codes for celiac disease rather than one. However, there was a loss of sensitivity when the criteria were more restrictive.[11]

In Sweden[12–14] and Italy,[15] celiac disease-specific incidence registers that cover a proportion of the population and include information on histology and other diagnostic criteria have been established.

A validation study of the biopsies registered in the Swedish database of pathology found that 90% of the biopsies were correctly classified as indicating villous atrophy and 56% correctly classified as indicating small intestinal inflammation. Among the 121 patients registered as having villous atrophy, 89% had celiac disease based on a review of their medical records.[17]

A Finnish study validated 10% of the cases of governmental reimbursement for gluten-free food, and found that 1.17% of the cases did not fulfill the diagnostic criteria, despite the fact that this was a prerequisite for receiving the reimbursements.

Other studies have used national patient registers[19] or discharge registers[20] without[16] carrying out a validation of the diagnoses.

Current diagnostic guidelines provide a gold standard for diagnosis, but diagnoses rely on the validity and completeness of the data on the items in the guidelines. In everyday clinical practice, these data may be of low quality, missing, or misinterpreted, which leads to inferior diagnoses (compared to gold-standard diagnoses). This study reflects clinical practice, as it involves data that are inherently somewhat unclear and incomplete compared to ideal data.

We accessed medical records for 95% of the children that were registered as having celiac disease in the Danish National Patient Register. These medical records contained data on celiac disease-specific antibody-test results, but some results may have been missing. The oldest results are more likely to have been missing because they were not computerized. These older results can be more important than the more recent results, because they represent the tests that were completed at the time of each patient's first diagnosis. We sought to diminish the issue of missing data by obtaining information from the two primary laboratories that carried out the relevant tests during the early study period. In addition, early on in the study period, in accordance with the guidelines, all the children should have been diagnosed on the basis of a biopsy, and thus all the diagnoses should have been included in the Patobank. A negligible number of patients refused to have a biopsy, as a biopsy was the only recommended confirmatory test and having a biopsy was necessary in order for them to be financially reimbursed for gluten-free food.

The current diagnostic guidelines are based on tests with high sensitivity and specificity, but we do not have information on the manufacturers of the test kits used during the study period, which would have allowed us to check the sensitivity and specificity of the tests. However, although we have no information on whether the tests used would have met the criteria, we have no reasons to suspect that they would have failed to meet the criteria.

For the children who were not registered in the Danish National Patient Register, we only have information on antibodies from two laboratories. Therefore, the estimates of the completeness of the records in the Danish National Patient Register were underestimated and calculations of sensitivity and specificity were not possible. However, it was found that 20 children who had records of only tentative diagnoses in the Danish National Patient Register were categorized as having a definite diagnosis of celiac disease. The completeness of the Danish National Patient Register is an important factor for conclusions to be made about the prevalence of diagnosed celiac disease in Denmark, but it may be of less importance in epidemiological association studies of celiac disease, because up to nine in ten of those with celiac disease in Denmark are expected to be undiagnosed, and thus they are not registered.[21]

With regard to evaluation of the duodenal biopsies, the categorization according to the Marsh classification system relied on the pathologists' reports. Incorrect classifications may have occurred as a result of poor-quality biopsies or biopsies that failed to encompass the diseased areas of the duodenum, or some of the pathologists may have been inexperienced, leading to incorrect descriptions being recorded.[22–24] These

issues were not taken into account in this study unless they were specifically mentioned in the pathology reports. Some studies have reported a considerable degree of interobserver variation in SNOMED coding,[24–26] although other studies have reported that there is a good level of agreement between duodenal biopsies, SNOMED codes, and clinical diagnoses.[17]

Overall, our results may have been affected by uncertainties regarding the diagnostic items, the lack of information on symptoms, and the lack of other potentially relevant information. Incomplete information on biopsies or antibodies may have led to an underestimation of the number of correct registrations. However, our use of diagnostic criteria that are less strict than those recommended in the diagnostic guidelines may have led to an overestimation of the number of correct registrations.

The validity of the registrations seemed to have increased over time. This may be attributable to an increase in the number of correct registrations by clinicians due to improved diagnoses and/or a decrease in the number of missing test results over time. PPV was slightly increased when the diagnostic criteria associated with the registrations were made more restrictive by stipulating that there had to be two or more registrations. However, this occurred at the cost of missing multiple definite diagnoses of celiac disease. The considerable drop in the proportion of children diagnosed on the basis of biopsies after the implementation of the new diagnostic guidelines in 2013 confirms that validation should involve more than just pathology reports on biopsies.

The prevalence of celiac disease is increasing,[27,28] and there is substantial variation between countries.[21,29,30] This indicates that there are modifiable environmental factors that influence the development of the disease. Epidemiologic studies of the etiology of celiac disease are urgently needed if we are to identify the potentially preventable factors that are responsible for the increase in prevalence.[31] Studies based on data obtained from national register-based studies are useful, as they ensure that sufficiently large study populations can be studied and that the study populations are not self-selected. Databases based on accurate diagnoses are needed in order to use the data to address the key questions.

Our findings suggest that the Danish National Patient Register could be a valuable source to identify diagnosed cases of celiac disease. However, the validity of the celiac disease registrations is suboptimal, and validation of the data is warranted before they are used for research purposes.

Acknowledgments

Disclosure

The authors report no conflicts of interest in this work.

References

1. Giersiepen K, Lelgemann M, Stuhldreher N, et al. Accuracy of diagnostic antibody tests for coeliac disease in children: summary of an evidence report. *J Pediatr Gastroenterol Nutr*. 2012;54(2):229–241.
2. Husby S, Koletzko S, Korponay-Szabó IR, et al. European Society for Pediatric Gastroenterology, Hepatology, and Nutrition guidelines for the diagnosis of coeliac disease. *J Pediatr Gastroenterol Nutr*. 2012;54(1):136–160.
3. 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(1):330–354.
4. [No authors listed]. Revised criteria for diagnosis of coeliac disease: report of working group of European Society of Paediatric Gastroenterology and Nutrition. *Arch Dis Child*. 1990;65(8):909–911.
5. Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. *Scand J Public Health*. 2011;39(7 Suppl):30–33.
6. Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol*. 2015;7: 449–490.
7. Dydensborg S, Toftedal P, Biaggi M, Lillevang ST, Hansen DG, Husby S. Increasing prevalence of coeliac disease in Denmark: a linkage study combining national registries. *Acta Paediatr*. 2012;101(2):179–184.
8. Størdal K, Bakken IJ, Surén P, Stene LC. Epidemiology of coeliac disease and comorbidity in Norwegian children. *J Pediatr Gastroenterol Nutr*. 2013;57(4):467–471.
9. Emilsson L, Magnus MC, Størdal K. Perinatal risk factors for development of celiac disease in children, based on the prospective Norwegian Mother and Child Cohort Study. *Clin Gastroenterol Hepatol*. 2015;13(5):921–927.
10. Canova C, Zabeo V, Pitter G, et al. Association of maternal education, early infections, and antibiotic use with celiac disease: a population-based birth cohort study in northeastern Italy. *Am J Epidemiol*. 2014;180(1):76–85.
11. West J, Fleming KM, Tata LJ, Card TR, Crooks CJ. Incidence and prevalence of celiac disease and dermatitis herpetiformis in the UK over two decades: population-based study. *Am J Gastroenterol*. 2014;109(5):757–768.
12. Ivarsson A, Persson LA, Nyström L, et al. Epidemic of coeliac disease in Swedish children. *Acta Paediatr*. 2000;89(2):165–171.
13. Namatovu F, Olsson C, Lindkvist M, et al. Maternal and perinatal conditions and the risk of developing celiac disease during childhood. *BMC Pediatr*. 2016;16:77.
14. Ivarsson A, Hernell O, Nyström L, Persson LA. Children born in the summer have increased risk for coeliac disease. *J Epidemiol Community Health*. 2003;57(1):36–39.
15. Zingone F, West J, Auricchio R, et al. Incidence and distribution of coeliac disease in Campania (Italy): 2011-2013. *United European Gastroenterol J*. 2015;3(2):182–189.
16. Collin P, Huhtala H, Virta L, Kekkonen L, Reunala T. Diagnosis of celiac disease in clinical practice: physician's alertness to the condition essential. *J Clin Gastroenterol*. 2007;41(2):152–156.
17. 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;9:19.

18. Pedersen CB. The Danish Civil Registration System. *Scand J Public Health*. 2011;39(7 Suppl):22–25.

19. Sevelsted A, Stokholm J, Bønnelykke K, Bisgaard H. Cesarean section and chronic immune disorders. *Pediatrics*. 2015;135(1):e92–e98.

20. Sandberg-Bennich S, Dahlquist G, Källén B. Coeliac disease is associated with intrauterine growth and neonatal infections. *Acta Paediatr*. 2002;91(1):30–33.

21. Horwitz A, Skaaby T, Karhus LL, et al. Screening for celiac disease in Danish adults. *Scand J Gastroenterol*. 2015;50(7):824–831.

22. Corazza GR, Villanacci V, Zambelli C, et al. Comparison of the interobserver reproducibility with different histologic criteria used in celiac disease. *Clin Gastroenterol Hepatol*. 2007;5(7):838–843.

23. Mubarak A, Nikkels P, Houwen R, Ten Kate F. Reproducibility of the histological diagnosis of celiac disease. *Scand J Gastroenterol*. 2011;46(9):1065–1073.

24. Arguelles-Grande C, Tennyson CA, Lewis SK, Green PH, Bhagat G. Variability in small bowel histopathology reporting between different pathology practice settings: impact on the diagnosis of coeliac disease. *J Clin Pathol*. 2012;65(3):242–247.

25. Taavela J, Koskinen O, Huhtala H, et al. Validation of morphometric analyses of small-intestinal biopsy readouts in celiac disease. *PLoS One*. 2013;8(10):e76163.

26. Weile B, Hansen BF, Hägerstrand I, Hansen JP, Krasilnikoff PA. Interobserver variation in diagnosing coeliac disease: a joint study by Danish and Swedish pathologists. *APMIS*. 2000;108(5):380–384.

27. Lohi S, Mustalahti K, Kaukinen K, et al. Increasing prevalence of coeliac disease over time. *Aliment Pharmacol Ther*. 2007;26(9):1217–1225.

28. Ludvigsson JF, Rubio-Tapia A, van Dyke CT, et al. Increasing incidence of celiac disease in a North American population. *Am J Gastroenterol*. 2013;108(5):818–824.

29. Myleus A, Ivarsson A, Webb C, et al. Celiac disease revealed in 3% of Swedish 12-year-olds born during an epidemic. *J Pediatr Gastroenterol Nutr*. 2009;49(2):170–176.

30. Rubio-Tapia A, Ludvigsson JF, Brantner TL, Murray JA, Everhart JE. The prevalence of celiac disease in the United States. *Am J Gastroenterol*. 2012;107(10):1538–1544.

31. Ludvigsson JF, Green PH. The missing environmental factor in celiac disease. *N Engl J Med*. 2014;371(14):1341–1343.

Supplementary materials

Table S1 Celiac disease-related SNOMED codes

Topographic codes	Morphologic codes	Diagnostic codes
T63910, T63911, T63930, T63950, T63951, T64000, T64020, T64040, T64300, T64320, T64910, T63950, T65100, T65110, T65930, T65935	M42100, M43000, M43005, M43009, M43900, M47195, M58000, M58015, M58018, M71000, M72000, M72115	S62180

Note: We applied for Patobank records that contained relevant topographic and morphologic codes or diagnostic codes.
Abbreviation: SNOMED, Systematized Nomenclature of Medicine.

Table S2 Classification of duodenal biopsy procedures recorded in the Patobank, based on microscopic histological descriptions, conclusions, and selected SNOMED codes

SNOMED code	Classification of biopsies				
	Compatible with Marsh 2–3 (n=1,233)	Compatible with Marsh 1 (n=220)	Unclassifiable (n=44)	Compatible with Marsh 0 (n=1,078)	Total (n=2575)
S62180 (celiac disease)	697 (29)	48 (20)	2 (0)	3 (0)	750 (49)
M58018 (crypt hypertrophic villous atrophy)	824 (42)	3 (0)	9 (4)	7 (1)	843 (47)
M47195 (intraepithelial lymphocytosis)	261 (11)	124 (66)	4 (1)	23 (16)	412 (16)
M4 (inflammation and fibrosis)	717 (30)	217 (94)	14 (10)	715 (590)	1673 (724)
M5 (degeneration, necrosis, deposition, atrophy)	1027 (52)	11 (2)	19 (6)	57 (43)	1114 (103)
M4 or M5 or S62180	1232 (63)	219 (94)	34 (11)	736 (603)	2221 (771)

Note: Data shown as number (number not registered in the Danish National Patient Register).
Abbreviation: SNOMED, Systematized Nomenclature of Medicine.

Table S3 Records of celiac disease-specific antibodies by the type of hospital department that provided the records[a]

Charateristics	Medical records accessed			Medical records not accessed	Total
	Pediatric departments	Departments of internal medicine	Other departments		
Departments[c]	17	22	7	26	72
Children[c;d]	1,927; 104 (25–276)	178; 4 (1–42)	30; 2 (1–17)	112; 2 (1–28)	2,247; 7 (1–276)
Antibody registrations	10,021	1,111	72	48[b]	11,252
Antibody registrations per child[d]	5 (0–30)	6 (0–27)	2 (0–7)	(0–7)[b]	4 (0–30)
≥1 positive antibody tests[e]	1,346 (70)	143 (80)	12 (40)	9[b]	1,510 (67)
tTG2 at ≥10× ULN or positive EMA[e]	998 (52)	109 (61)	7 (23)	6[b]	1,120 (50)
Biopsy compatible with Marsh 2–3[e]	955 (50)	121 (68)	8 (27)	43 (38)	1,127 (50)

Notes: [a]2,247 of the children were registered as having celiac disease in the Danish National Patient Register; [b]results that were based solely on antibody-test results from the SSI laboratory or the GastroLab (numbers expected to be higher when medical records accessed); Data shown as [c]number; [d]median (range) or [e]number (%).
Abbreviations: ULN, upper limit of normal; EMA, endomysial antibody; HLA, human leukocyte antigen; SSI, Statens Serum Institut; tTG2, anti-tissue transglutaminase 2.